

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



ONE YEAR. TEN STORIES.

The newsmakers of the year **PAGE 507**

NEWS REVIEW

THAT WAS A YEAR, THAT WAS

2016 was one damn
thing after another...

PAGE 496

POLICYMAKING

SPEAK UP FOR SCIENCE

It's more important than
ever to build public trust

PAGES 517 & 520

CALENDAR

EVENTS DIRECTORY 2017

The Nature guide to global
science events and courses

BACK PAGES & NATURE.COM

NATURE.COM/NATURE

22/29 December 2016 • \$10

Vol. 540, No. 7634



THIS WEEK

EDITORIALS

GAMBLING Seasonal odds reflect more than statistics **p.484**

WORLD VIEW Scientists under fire should join resistance **p.485**



TRIMMINGS Pots show early humans ate their greens **p.487**

Europe should retain its ambition

The EU is planning a new round of billion-euro Flagship research projects, learning from its rocky experience so far. It must be open to all types of science.

Despite the rising tide of Euroscepticism, exemplified most obviously by this year's Brexit vote, the European Commission still has big ambitions for the continent's science. Preparations for the EU's ninth framework research programme (FP9), to be launched in 2021, will begin in earnest next year. And, surprisingly perhaps given recent history, it will include another couple of ten-year, billion-euro FET Flagship projects.

FET stands for future and emerging technologies, and its flagship projects aim to promote digital technologies while addressing policy priorities. In commission-speak, they are supposed to be 'game-changing' and 'visionary', with the potential to deliver 'transformational economic or social impact'. These are lofty but laudable goals that deserve support. But, given their immense cost, the projects must be carried out in the best way. And the commission is still trying to work out what that is.

To say that the concept got off to a rocky start would be an understatement. The Human Brain Project, one of the two initial flagships selected in 2013, bitterly divided the European neuroscience community and had to be restructured after a painful and public showdown. And although the second, called Graphene, has trundled along peacefully, many scientists and officials across Europe wondered if the commission would risk launching any more. In fact, it quietly launched a third — a quantum-science project — earlier this year, without the razzmatazz of the public competition that accompanied the first round. 'Quantum' is considered important to the development of the European Open Science Cloud, a virtual environment for storing and sharing data, and its various security-related digital ambitions. It is due to launch in 2018 (see go.nature.com/1semikt).

STEERING THE FLAGSHIPS

On 15 December, member states and European research organizations met in Brussels to discuss what comes next. The meeting followed the outcome of a public consultation in which scientists were invited to submit possible flagship ideas. In the next few months the commission will settle on three or four themes; specific calls for proposals will be issued in mid-2017.

Last week's meeting settled on three general themes, which can still change or expand. One is energy, environment and climate change; another is health and life sciences; and the third, described as 'ICT for the connected society', covers communications technology. Then, as in the first round of flagships, five or six winners will be given a generous grant to develop their concepts, and one or two of these will be selected for funding within the next framework programme.

Of the two dozen submissions already sent in by scientists, the commission felt that 14 ticked the requisite boxes of size and cross-nationality, ambition, feasibility and alignment with EU policy priorities. The most popular ideas — and these offer early clues to how the commission will frame its official call for proposals — were in

health care, climate change and food security, robotics and renewable energy. No surprises there. But other persuasive concepts ranged from synthetic biology and regenerative medicine to digital humanities.

How to choose between them? In post-Brexit, post-Human Brain Project Brussels, the mood is conciliatory towards member states. Not always fairly, the commission, an executive body, is often accused of making decisions without appropriate input from national politicians and external experts. Certainly, in the creation of its first two flagships, the commission gave academic-led consortia too much leeway to define their own management structures. This allowed the Human Brain Project temporarily to veer off course, concentrating too much power in too few hands and sidelining most of the project scientists. This time,

"The European Commission has been at pains to consult as widely as possible."

the commission has been at pains to consult as widely as possible with the scientific community, with research bodies and with industry — and most particularly with member states.

The two suggested projects with the most support so far are 'Future of Healthcare' and 'Robot Companions for Citizens', strong proposals that were both finalists in the first flagship round. Future of Healthcare is a consortium led by scientists in Berlin and Lausanne, Switzerland. It aims to develop the technologies and the legal and regulatory environment for a continent-wide digital health-care system, as well as the integrated molecular and imaging technologies that will help promote the development of personalized medicine. Robot Companions is led by scientists in Genoa, Italy, and will exploit multiple disciplines, from artificial intelligence to cognitive sciences, to develop soft-bodied, 'perceptive' robots for those who are old and lonely; it will also design robots that can help in surgical procedures, farming and other areas.

Some at last week's meeting said that scientists involved in related consortia should pitch a joint idea. That is not a good idea. As UK representatives warned, doing so could blur goals and create consortia too large to manage projects efficiently. Alas, the views of the United Kingdom are of diminishing importance in Europe since the Brexit vote. But the commission should think carefully about forcing such marriages — which would also prevent groups from competing against each other.

It should also assess the extent to which it will unquestioningly follow the views of the member states — whose national interests sometimes conflict with the broad European goals they have signed up to. Governments will, of course, lobby for money to be spent on health care and robots, which yield tangible and vote-winning outcomes. The case for massive resources for cross-national digital humanities projects — which aim to preserve shared history and culture — is a harder sell. That is why the commission must explicitly make room for them to compete when it sends out its funding call. Cooperation and shared values may be falling out of political fashion, but they remain the bedrock of the European project. ■

Snow blind

Have a bet on a white Christmas, but don't fall for an old chestnut.

It just took some treetops to glisten and some children to listen for Bing Crosby to enjoy a white Christmas. Bookmakers are a more cynical bunch, so usually demand to see at least a single white flake fall during the 24 hours of 25 December. The chances of snow at *Nature* HQ this weekend are diminishing as Britain basks in unseasonably mild conditions — bookmakers put the odds of an official white Christmas in London at about 10 to 1.

If that seems too long a shot, then science offers a way to make the bet more attractive to punters. By combining it with a second bet on an event much more likely to happen, bookies can exploit a psychological tic called the conjunction fallacy. Odds, for example, of a cover of the Rolling Stones' classic 'You can't always get what you want' reaching the UK Christmas number-one slot are a much shorter 9/4 — it's one of the favourites. (All odds correct as *Nature* went to press.) And although logic and statistics tell us that the chance of both events occurring must be lower than either of the single events alone, gamblers routinely fail to recognize that.

Study after study shows that pairing with a dead-cert makes an unlikely wager seem more — not less — likely to happen. And that makes people more willing to put money on an outsider. This logical illusion can explain much fixed-odds betting on sport, including football. Gamblers routinely think there is more chance that West Bromwich Albion will win at Arsenal on Boxing Day (9/1) if the wager is combined with a Manchester United home victory over Sunderland (2/9).

Exactly why this happens is not clear, but it seems that some gamblers play the odds off against each other in their heads, and assume — incorrectly — that the combined chance of the two is an average of the odds, that the extreme likelihood of the second option somehow tempers the outlandishness of the first. It can be an expensive mistake.

With supreme knowledge of the human condition, one might think that scientists would be immune from making rash bets. Not so. This

year, astrophysicist Shrinivas Kulkarni has lost a US\$1,000 wager on the origins of fast radio bursts, and another astrophysicist, David Wiltshire, has stumped up \$200 for a lamp after losing a 10-year wager with a colleague on the role of the cosmological constant in dark energy.

Although such simple bets between researchers (sometimes friendly and sometimes not so) are a long-standing feature of science, perhaps the most lucrative are those in which scientists (just like bookmakers) pit their calculated professionalism against the optimism and emotion of those who follow a lost cause. This year has also seen climate scientists cash in on bets made with sceptics about the continued warming of the planet. Indeed, the annual meeting of the American

“Bookies can exploit a psychological tic called the conjunction fallacy.”

Geophysical Union last week had a session dedicated to betting on climate change.

Bets accepted or refused can be a good way to gauge how firmly a sceptic truly believes their contrarian position, because wagers typically follow strongly and honestly held (however unlikely) opinions. (Hence, some fans of Sunderland will see the odds of 14/1

on them winning the above match as too good to turn down.)

Some events baffle punters, scientists and bookmakers alike — and 2016 has seen plenty of those. So who would dare to argue that a theoretical 'social-physics' model — used, among other things, to predict the behaviour of plastic crystals — would do a worse job than pollsters and experts at predicting the results of political votes such as this year's Brexit referendum and US presidential election? Physicists last month published such a mean-field model, which they say describes the dynamics of two-group conflicts on the basis of the interactions between group members, opponents and how willing people are to change their minds (H. T. Diep *et al. Physica A* **469**, 183–199; 2017).

The model's output shows whether each side in a political dispute will tend towards negotiation or conflict, and the often wild swings and oscillations in their attitudes towards each outcome along the way. It is not a tool of prediction, the physicists caution, but rather one of anticipation for strategic purposes. That seems a sensible approach given recent events, which have shaken faith in predictions of all sorts. So in that spirit, as 2016 draws to an end and as Bing almost sang: your days may be merry and bright, and all your Christmases may be white. ■

On retirement

When great colleagues end their careers, employers should recognize their value.

According to the US psychologist David Hershenson, people who retire can experience up to six separate statuses. Retrenchment comes when they cut back on work and principal employment, and Exploration sees them think about what activities to do next instead. In their Try-out status, retirees see how well suited they are to new activities (including inactivity), and Involvement marks their long-term participation in pursuits they enjoy and can stick with. When new options present themselves, retirees are faced with Reconsideration. And should they move on from an activity, or indeed return to work, then they Exit. Not coincidentally, the six statuses together form the acronym RETIRE (D. B. Hershenson *J. Aging Studies* **38**, 1–5; 2016).

It may seem contrived, but the study of retirement — and finding ways to investigate it — is an important business as the population ages. Not least is the question of who should be paid to retire, by whom and when. As funds dwindle, retirement ages are creeping up. But

do some workers deserve an earlier break from the daily grind than others? The government of the Netherlands has put some serious thought into whether people in some professions — particularly occupations involving heavy manual labour, such as construction — should have their retirement age fixed or reduced, even while people in less-demanding jobs see their retirement ages rise (N. Vermeer *et al. Labour Econ.* **43**, 159–170; 2016). In the United Kingdom, the opposition Labour Party leader Jeremy Corbyn has suggested something similar. As Jane Austen wrote in *Sense and Sensibility*: “It isn't what we say or think that defines us, but what we do.”

Policies on retirement, then, and the studies that inform them, need to broaden their assessment to include an earlier status: work. If retirement is a well-earned break after a long and productive career, how can researchers distinguish those employees who should enter it before some and after others?

At *Nature* we have our own internal scoring system, with its own (slightly) contrived acronym. We look for people who perform Work that is consistently Excellent and Notable, and that has helped to define the cultural and scientific Zeitgeist for a significant time — usually measured at about 40 years. We call it the WENZ measure. Every organization should have a WENZ. And when people with the WENZ factor retire, they should do so in the full and certain knowledge that their contribution has been valued by colleagues. More than that, they should know they will be missed. ■



How Woody Guthrie can help us fight for science

After the election of Donald Trump, Jacqueline M. Vadjunec offers a message of resistance and hope from deep within the US Bible Belt.

When I moved from Massachusetts almost a decade ago to teach at Oklahoma State University, many colleagues were afraid for my career. I work on the human dimensions of global environmental change, and Oklahoma has a long and complex history with science, including climate change.

Oklahoma was the first state to ratify 'anti-Darwin' legislation in 1923 and today is home to key sceptics in the war on climate change, including Republican Senator James Inhofe and Scott Pruitt, the state's attorney-general, who earlier this month was nominated to run the US Environmental Protection Agency. These politicized debates trickle down, and both evolution and human-induced climate change remain contested topics, especially in schools.

However, Oklahoma is also the home of protest singer Woody Guthrie, a visible example of resistance in the 1930s class and culture wars between rural and urban values. If Woody could use his voice to speak up, so can scientists.

In truth, my career is fine, and my colleagues are supportive. I not only manage, but also thrive. And if I can, then so can other scientists who find themselves concerned about the tidal wave of climate scepticism that comes with last month's election of Donald Trump and his associates. The election might have powerful effects on science, policy and funding. But I want to stress the power and promise of human agency.

In my case, adjustments are minor, but might seem substantial elsewhere. I realize that in my day-to-day actions in the classroom and in my research with family farmers and ranchers, I probably hold a minority viewpoint on human-induced climate change. In the classroom, I am sensitive to the fact that many of my students have family ties to the oil and gas industry. I regularly see them struggle with the local contradictions. I try to create a place of mutual respect to embrace this struggle on their own terms, while also trying to focus on our role as global citizens facing global challenges. It is not always an easy balancing act; these experiences have taught me that most students care about global environmental change, but often have little previous exposure to such issues — in part because of the decisions of local politicians and school boards. In our debriefing at the end of the semester, students often express frustration that they weren't exposed to many of the issues surrounding climate change at a younger age.

I also learned that actively listening to (instead of talking at) farmers and ranchers who care about sustaining their land and livelihoods is a good way to open dialogue. We can then find common ground on pressing environmental issues, such as the depletion of the Ogallala Aquifer, encroachment of invasive and nuisance woody-plant species on pasture lands, and the compounding impacts of long-term cyclical

drought. People in Oklahoma care about the long-term sustainability of their natural resources, but they often use language that is different from that of climate scientists and elected officials.

We should remember the power of the small. In *Weapons of the Weak* (Yale Univ. Press, 1985), James C. Scott illustrates the power of "everyday forms of resistance". It is through these small acts (both intentional and unintentional) that power can be contested, destabilized and renegotiated. There may be increased climate scepticism, but there will also be more scientists, teachers and citizens banding together to respond.

Despite official policies that limit climate-change education, a recent survey of 115 science teachers in Oklahoma showed that more than 80% teach climate change in state schools, either formally or informally (N. M. Colston and T. A. Ivey *J. Educ. Policy* **30**, 773–795; 2015). Faced

with few locally available teaching resources, most teachers write their own lesson plans. They also take advantage of 'teach the controversy' campaigns — intended by some to undermine the scientific consensus — to introduce students to locally controversial topics such as human-induced climate change, which otherwise might be seen as off limits. Consequently, more than two-thirds of these teachers say that they experience no pushback from students, parents or administrators (N. M. Colston and J. M. Vadjunec *Geoforum* **65**, 255–265; 2015).

In resisting the mood of anti-science, researchers need to reach out to a diverse public in more accessible ways. We also need to accept different ways of knowing or even talking about climate change: ways that open doors to start a conversation;

ways that are more context specific, culturally sensitive and nuanced than science in general might be comfortable with.

For example, state politics in Oklahoma are shaped in part by continuous interactions with Oklahoma's 39 Tribal Nations. These Native Americans, in conjunction with researchers and media artists, are speaking up to provide their own unique perspectives on climate change (see go.nature.com/2grktji). Such projects show that indigenous people care deeply about climate-change issues, but that when it comes to adaptation and mitigation, they would like their traditional knowledge to be valued along with that produced by Western science.

Where there is climate war, there is also climate resistance, in large and small ways. I urge scientists not to lose heart, but to develop tools and projects that are useful to citizens, as well as to our peers and funding agencies. In practising such open-minded science, we might find that we have more allies than are visible at first glance. ■

Jacqueline M. Vadjunec is an associate professor in the Department of Geography at Oklahoma State University in Stillwater.
e-mail: jacqueline.vadjunec@okstate.edu

WHERE THERE IS
CLIMATE WAR,
THERE IS ALSO
CLIMATE
RESISTANCE,
IN LARGE AND
SMALL WAYS.

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

EVOLUTION

Penis bone lost through evolution

Our monogamous lifestyle may explain why humans, unlike many other mammals, lack a penis bone.

The bone, called a baculum, rests at the end of the penis and is thought to provide structural support and prolong copulation. Matilda Brindle and Christopher Opie at University College London analysed the size of bacula in nearly 2,000 species of mammal, including primates and carnivores. They found that species that copulate for longer tend to have longer bacula. The same is true of animals that have more than one mate or have seasonal-breeding patterns, which lead to intense competition between sperm from different males after mating.

The results show that the baculum first evolved 145–95 million years ago, in the common ancestor of primates and carnivores. It disappeared from the human lineage after our split with chimpanzees, and this may have coincided with the switch towards a more monogamous lifestyle, the authors say.

Proc. R. Soc. B 283, 20161736 (2016)

APPLIED PHYSICS

Device breaks cooling record

A cooling technology, if scaled up, could decrease temperatures by as much as 37°C — potentially boosting the capabilities of refrigeration equipment.

Shanhui Fan at Stanford University in California and his team built a device that includes a thermal emitter that gives off heat in the

mid-infrared range. Such wavelengths correspond to the 'transparency window' of Earth's atmosphere, allowing the heat radiation emitted by the apparatus to be released into space. Previous systems have reduced their temperatures by up to only 20°C in low-altitude areas with moderate humidity levels. But the team broke this record with the addition of several features, including a vacuum chamber around the emitter. This ensured that heat was emitted only into space, and not into the surrounding air, increasing the amount

of heat removed.

With further development, the technology could be used for applications including energy-efficient air conditioning.

Nature Commun. 7, 13729 (2016)

AGEING

Genes that make mice youthful

Four genes that reprogram adult cells into embryonic-like stem cells can also reverse some signs of ageing.

The four genes encode Yamanaka factors, which

monsoon patterns might shift by the end of the century.

The models suggest that monsoon rain will generally become more intense in a warming world. But the projected increase was 30% smaller, on average, when the team accounted for changes in land use and land cover. Shifts in monsoon rainfall might affect regional water resources and agricultural yields, the authors say. *Geophys. Res. Lett.* <http://doi.org/bv38> (2016)



DAVID SANTIAGO GARCIA/ALAMY

CLIMATE SCIENCE

Smaller monsoon boost predicted

Climate change may produce smaller-than-expected increases in rainfall in the world's monsoon regions over the coming decades, thanks to changes in land use.

More than 70% of the global population live in monsoon areas. Benjamin Quesada of the Karlsruhe Institute of Technology in Germany and his colleagues ran global climate models with and without projected deforestation and other land-use changes to compare how

are essential for embryonic development, but usually cause tumours when expressed long-term in animals. Juan Carlos Izpisua Belmonte at the Salk Institute in La Jolla, California, and his colleagues switched the genes on for two days per week over several weeks in mice that had an ageing disorder called progeria. The animals lived about 30% longer, and showed improvements in tissue healing and other signs of ageing, such as organ failure. In normal aged mice, switching on the genes led to improved recovery from

muscle injury and to other signs of youthfulness. The mice did not develop cancer.

The authors link the rejuvenation to epigenetic remodelling — changes in the chemical marks on DNA that do not alter its sequence but influence gene expression.

Cell 167, 1719–1733 (2016)

HIV

Targeting host genes for therapy

By inactivating any one of five human genes, scientists can prevent HIV from entering and growing in immune cells.

Antiviral therapies targeting host genes that the virus depends on, rather than targeting the virus itself, are promising because these genes do not mutate as frequently as viruses do. This could avoid the development of drug resistance. Bruce Walker at the Ragon Institute of MGH, MIT and Harvard in Cambridge, Massachusetts, and his colleagues screened the genome of human T cells and identified five genes not essential to cell survival whose inactivation protected cells from HIV infection. Cultured cells lacking these genes resisted HIV infection. The genes encode proteins that facilitate virus entry into human cells, and one that mediates cell aggregation, which allows the virus to spread between cells.

The authors say their approach could also be used to find drug targets for other pandemic viruses.

Nature Genet. <http://dx.doi.org/10.1038/ng.3741> (2016)

OCEAN SCIENCE

East Antarctic ice melts from below

Ocean heat is melting a floating ice shelf in East Antarctica, raising concerns of accelerated glacier discharge and sea-level rise.

East Antarctica's large ice sheet was thought to be more stable than that of

West Antarctica. But Stephen Rintoul at the University of Tasmania in Hobart and his colleagues found that increasing ocean heat is weakening the Totten Ice Shelf in East Antarctica. The shelf supports glaciers whose volume is equivalent to 3.5 metres of global sea-level rise.

The team analysed oceanographic data collected in 2015 and found deep channels at the front of the ice shelf through which large volumes of temperate deep-ocean water flow into the ice shelf's cavity. *Sci. Adv.* 2, e1601610 (2016)

ARCHAEOLOGY

Early humans cooked vegetables

Humans cooked and ate a variety of plants — mostly grasses and aquatic plants — as early as 8,200 years ago.

Scientists have often found signs of prehistoric meat and milk processing, but direct evidence of early plant cooking has been rare. Richard Evershed at the University of Bristol, UK, and his colleagues analysed residues from 110 ancient pottery fragments discovered in the Libyan Sahara. The pieces were found in a cave and a rock shelter, both of which also housed well-preserved plants several thousand years old (legumes of *Cassia* pictured). Besides animal fats, the team found plant lipids in most of the pots.

Some pots seem to have been used exclusively for fruits and seeds, but the team also uncovered evidence of leaves and stems being cooked.

Nature Plants <http://go.nature.com/2gns862> (2016)



CANCER GENETICS

Why melanoma is worse in men

Differences in the expression of a particular gene could explain why men with skin cancer tend to have a lower survival rate than women.

The gene, *PPP2R3B*, is expressed from both X chromosomes in women and from the X and Y chromosomes in men. Alan Spatz at McGill University in Montreal, Canada, and his colleagues studied tissue samples from people with melanoma, and found that greater expression of *PPP2R3B* correlated with longer survival times. In cultured cells, high levels of *PPP2R3B* expression slowed melanoma growth by interfering with DNA replication and slowing cell division.

Expression of the gene was higher in women than men, which could explain why women with melanoma have better clinical outcomes, the authors say.

Sci. Transl. Med. 8, 369ra177 (2016)

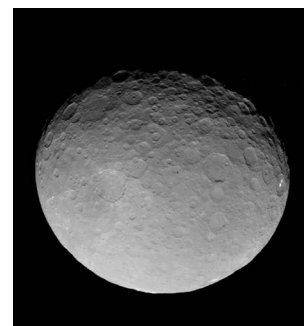
PLANETARY SCIENCE

Where Ceres hides its water

Frozen water has been lurking beneath the rocky surface of the Solar System's biggest asteroid since its birth billions of years ago.

NASA's Dawn spacecraft began orbiting Ceres (pictured), which is also a dwarf planet, in 2015. This allowed a team led by Thomas Prettyman at the Planetary Science Institute in Tucson, Arizona, to measure hydrogen at the asteroid's surface. Water inside Ceres chemically alters the surface, leaving a hydrogen imprint there. The highest hydrogen concentrations appeared at mid to high latitudes.

A second study looked at ice trapped in permanently shadowed regions of Ceres' surface. Of the 634 craters



studied, only 10 contained icy material, say Thomas Platz of the Max Planck Institute for Solar System Research in Göttingen, Germany, and his colleagues. Ceres, like Mercury and the Moon, can apparently trap frozen water in dark areas for long periods of time, they add.

Science <http://doi.org/bv3z>; *Nature Astron.* 1, 0007 (2016)

GENOME EDITING

Enzyme switches turn CRISPR off

Inhibitors of a gene-editing system called CRISPR–Cas9 could one day provide a safety switch, allowing researchers to halt the system's activity in cells.

CRISPR–Cas9 is a naturally occurring bacterial defence mechanism that scientists have harnessed to alter DNA sequences. Alan Davidson at the University of Toronto in Canada and his colleagues searched for bacterial proteins that block the DNA-slicing Cas9 enzyme from the bacterium *Neisseria meningitidis*, and found three families of such proteins.

The inhibitors halted CRISPR–Cas9 editing in human cells, suggesting that they could be used to better control genome editing. They could be important if CRISPR–Cas9 is to be used for gene therapy in people, or to edit the genomes of entire populations in the wild.

Cell 167, 1829–1838 (2016)

➔ **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch

SEVEN DAYS

The news in brief

POLICY

Gene triangle

Britain's fertility regulator has decided to allow, in "certain, specific cases", the birth of babies from embryos that have been modified to contain three people's DNA. On 15 December, the UK Human Fertilisation and Embryology Authority announced that clinics can start to apply for licences to conduct limited trials of the technique, which aims to prevent mothers from passing on mutations in cellular organelles called mitochondria. The move makes the United Kingdom the first country to explicitly permit the controversial therapy.

'Corrosive' Brexit

Uncertainty in the wake of Britain's vote to leave the European Union is having a "corrosive effect" on UK science that could cause

SOUND BITE

“If Trump turns off the satellites, California will launch its own damn satellite.”

California governor Jerry Brown responds to suggestions that budget cuts could threaten Earth-observing-satellite programmes, at a meeting of the American Geophysical Union in San Francisco, California, on 14 December.



MEHDI FEDOUACH/AFP/GETTY

Replica of ice-age cave opens in France

A replica of Lascaux, a cave in southwestern France that is famous for its galleries of ice-age paintings, opened its doors to the public on 15 December. The original cave has been closed to visitors since 1963, after heavy tourist traffic caused the stunning paintings, estimated to be 18,000 years old, to deteriorate.

The €57-million (US\$59-million) centre, Lascaux 4, is at the foot of the hills in which the original was discovered in 1940 and is a replica of almost all of the cave, including its dark and damp atmosphere. The first replica of the cave, which opened in 1983, featured just the two main galleries.

FACILITIES

Mini accelerator

Physicists are a step closer to creating a miniature particle accelerator, it was announced on 14 December. The Advanced Wakefield Experiment, or AWAKE, based at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, has yet to accelerate particles. But tests in its initial week of operation showed for the first time that pulses of protons can generate the wave of plasma needed to do just that. Harnessing the effect, which had previously been seen only in simulations, could eventually lead to smaller, cheaper particle accelerators.

Telescope setback

A judge in Hawaii has overturned the 2014 state approval of the Thirty Meter Telescope (TMT) consortium's sublease with the University of Hawaii at Hilo, which the project needs to build its US\$1.5-billion instrument on Mauna Kea. Plans to build the telescope have been mired in conflict, but the 15 December ruling is a smaller stumbling block than the state supreme court's decision a year ago to rescind the building permit for the project, on which a fresh round of hearings is under way. The sublease ruling stems in part from a legal challenge from Native Hawaiians, some of whom say that the TMT will desecrate sacred land. The

telescope has a back-up site in the Canary Islands if it cannot be built in Hawaii.

PEOPLE

Science exile

In a display of solidarity with troubled particle physicist Adlène Hicheur, scientists held an international high-energy-physics workshop on 13 December in the small town of Vienne, southeast France, where Hicheur is under house arrest. Hicheur had previously been jailed in France for alleged terrorism offences — a conviction strongly disputed by him and his colleagues — and after his release in 2012 had restarted his research career in Brazil. He was mysteriously deported from Brazil in July. Having renounced his French nationality in October, the Franco-Algerian physicist intends to fly to Algeria within two weeks. French authorities have agreed to lift the house arrest on his departure day.

Trump energy pick

US president-elect Donald Trump nominated Rick Perry to run the US Department of Energy on 14 December. Perry (pictured) governed Texas from 2000 to 2015 and sought the Republican presidential nomination in 2012. As governor, he supported fossil-fuel production, and questioned the science



underlying climate change. Critics have voiced concern over his lack of scientific or technical background. In 2013, he proposed eliminating the energy department. Despite his ties to the fossil-fuel industry, the share of energy production from renewables in Texas increased substantially during his term as governor (see page 492). Trump has also reportedly selected Montana congressman Ryan Zinke to head the Department of the Interior, which oversees federal public lands, natural resources and Native American programmes. Like Perry, Zinke has expressed doubt over human-induced climate change. He has voted in favour of coal extraction and oil and gas drilling. Both nominations will need confirmation by the Senate.

No Stamina

The Republic of Georgia has banned controversial stem-cell entrepreneur Davide Vannoni from working in the

country. In March last year, Vannoni was convicted in Italy on charges of conspiracy and fraud for administering unproven stem-cell therapies in that country, but his sentence was suspended on the condition that he halt his procedures. In October, Italian prosecutors investigated allegations that his Stamina Foundation was offering treatments again, in Tbilisi. They sent documentation about Vannoni's case to the Georgian government, which responded with the ban, according to news reports.

EVENTS

Polar adventure

The Antarctic Circumnavigation Expedition (ACE) set off from Cape Town, South Africa, on 20 December on a three-month research cruise around the frozen continent. A 55-strong international research crew on board the *Akademik Treshnikov*, a Russian research vessel chartered for the voyage, will collect a variety of marine data for studies on the impact of climate change in the Southern Ocean. Swedish philanthropist Frederik Paulsen, founder of Ferring Pharmaceuticals, is the main sponsor of the expedition, which has been organized by the newly established Swiss Polar Institute in Lausanne.

NUMBER CRUNCH

55,006

The overall number of doctorate recipients in the United States in 2015, of whom 25,403 were female.

Source: National Science Foundation

Newton first edition

A rare copy of Isaac Newton's groundbreaking work *Philosophiæ Naturalis Principia Mathematica* has become one of the most expensive printed science books in history. On 14 December, an anonymous bidder paid US\$3.7 million for a first edition of the book at an auction at Christie's in New York City — more than twice as much as the auction house had expected. First published in 1687, the work includes Newton's law of universal gravitation and his laws of motion. A copy that had been presented to King James II of England sold for \$2.5 million in 2013.

Eczema activity

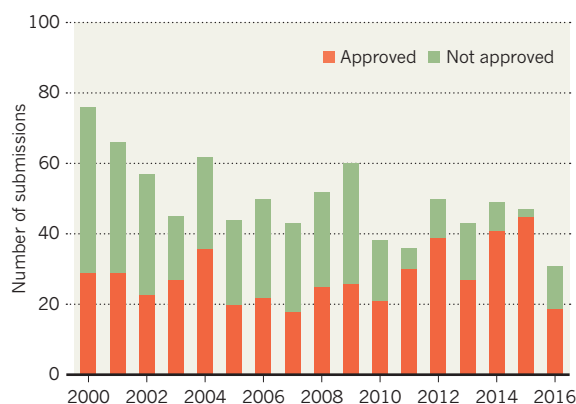
On 14 December, the US Food and Drug Administration approved the first new drug to treat eczema — a chronic inflammation of the skin that causes severe itching — in more than a decade. The drug, an ointment called Eucrisa (crisaborole), inhibits the protein phosphodiesterase 4 and was developed by Anacor Pharmaceuticals of Palo Alto, California. Two days later, Swiss drug-manufacturing giant Novartis announced its intention to buy a separate company, Ziarcro in Sandwich, UK, which is developing an oral eczema treatment that targets a histamine receptor. Novartis did not disclose how much it was paying for the company.

TREND WATCH

US drug approvals fell by more than 50% in 2016, according to a 14 December presentation by an official at the US Food and Drug Administration (FDA). The agency has approved 19 new drugs so far this year, its lowest annual tally in nearly a decade. The FDA attributed the decline to fewer submissions and the approval of five drugs ahead of schedule in 2015. The agency also rejected more drugs: in 2016, 61% of the FDA's decisions were approvals, compared with more than 95% in 2015.

STEEP DROP IN US DRUG APPROVALS

The US Food and Drug Administration approved only 19 new therapies in 2016, marking a 10-year low.



► NATURE.COM

For daily news updates see:

www.nature.com/news

NEWS IN FOCUS

POLITICS Scientists take steps to defend their fields under Trump **p.492**

NAVIGATION Why Galileo satellite fleet is a big deal for science **p.493**

2016 IN REVIEW The science events that shaped 2016 **p.496**



PROFILES Ten people who mattered in science this year **p.507**

US NAVY



Marine surveillance drones are likely to be among the technologies funded by a new EU research fund.

FUNDING

Peaceful EU starts to fund military research

Shift in focus comes in response to a changing world order and the threat of terrorism.

BY ELIZABETH GIBNEY

Faced with a changing world order and buffeted by a slew of political crises and terrorist attacks, the historically civilian European Union is bolstering its military capabilities. And that means making its first major investment in military research.

On 1 December, the European Parliament approved a €25-million (\$26-million) fund

dedicated to military research. It will form part of a proposed broader European Defence Fund, aimed at making military innovation more efficient and enlarging Europe's industrial defence base.

The research portion of the fund will cover electronics, advanced materials, encrypted software and robotics. The European Commission, the EU's policymaking arm, expects to invest a total of €90 million by 2020. It hopes the figure

will rise to €500 million a year for defence research from 2021. The sum is dwarfed by the EU's major research-funding programme, Horizon 2020, which will hand out €80 billion over 7 years, the €8.8 billion spent by EU member states on defence research in 2014, and what the United States and probably China spend on defence research (see 'Military metrics').

But some scientists fear that funding defence research is a step in the wrong direction for ▶

► the EU. “It will necessarily divert much-needed funding from civilian R&D budgets, at a time when they are urgently needed for areas such as climate and energy,” says Stuart Parkinson, executive director for UK-based advocacy group Scientists for Global Responsibility.

One of the EU’s main objectives is to promote peace. In the past, defence was seen as a national issue rather than something for the bloc to handle. The decision to create the research fund is in part driven by a drop in national defence-research funding, which declined by 18%, or €1.9 billion, between 2006 and 2014 in real terms, according to the European Defence Agency (EDA) in Brussels, which will manage the research fund on behalf of the commission.

The perception that international security is under threat is a driver for the broader defence fund. In November, the European Parliament passed a motion that says terrorists are targeting the continent on an unprecedented scale, and that Europe is “now compelled to react to an arc of increasingly complex crises”. The motion notes that for the first time since the Second World War, “borders in Europe have been changed by force” — referring to the annexation of Crimea in 2014 and the incursion into Ukraine by Russian forces.

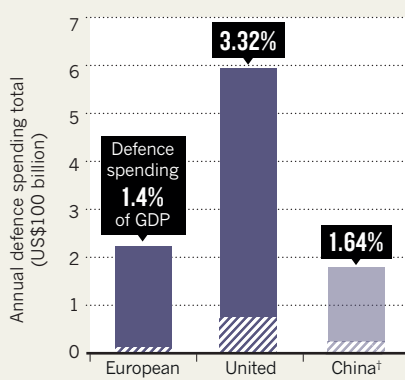
In September, Jean-Claude Juncker, president of the commission, made a similar point when speaking about the European Defence Fund. “Europe can no longer afford to piggy-back on the military might of others,” he said.

The rules for participating in the research fund are still being discussed, but it will be modelled loosely on Horizon 2020. It will probably promote projects that combine

MILITARY METRICS

The European Union’s 28 member states combined spend much less on defence research than the United States, and probably China.

▤ Spending on military R&D



*Figures for the EU from 2014, figures for the United States and China from 2015; †Figures are approximate based on an estimated minimum for defence total and the average of a range for percentage spent on research.

researchers from industry and academia and from different countries, says Denis Roger, a director in charge of research at the EDA.

But whereas researchers on Horizon 2020 projects are expected to publish their results, or to patent products for anyone to license, the commission is likely to restrict how the results of defence-fund research are publicized, classifying some and restricting licensing to national ministries. The EU has no army of its own — although Juncker has said he would like to create one. Instead, national ministries, alongside members of the defence industry, will be involved in setting priorities for the scheme, says

Roger. And unlike Horizon 2020, which welcomes participation from more than a dozen ‘associated’ countries, the research defence fund is likely to be open only to EU member states and Norway.

The project could boost certain fields. Roger says it will include research into metamaterials, which are made of tiny structures that manipulate the path of light and could potentially hide objects from radar, as well as methods of energy storage, flexible radio antennas that can be incorporated into clothing, and prototype maritime surveillance drones.

“I would imagine that a lot of countries would definitely see this as another opportunity for funding,” says Ortwin Hess, a physicist at Imperial College London. He notes that US scientists working on photonics and metamaterials can readily access defence funding. “My US colleagues wouldn’t survive without it. They live on it.”

Hess, who has received defence funding from the US and UK governments in the past, says he is a realist when it comes to what he calls the moral question. “I have to accept that our society has values that deserve to be defended,” he says. The military will adopt technologies developed in the civilian domain, and sometimes technology transfer can go the other way, he adds.

But Parkinson says that defence research often supports military efforts beyond actual defence, as well as the export of weapons to other countries. “Our view is that we need a much stronger focus on R&D which contributes to tackling the root causes of conflict — including a range of social and environmental problems.” ■

POLITICS

US scientists plan for an uncertain future

Concerned by the president-elect’s choice of advisers, researchers take steps to defend their fields.

BY JEFF TOLLEFSON AND ALEXANDRA WITZE
IN SAN FRANCISCO, CALIFORNIA

Incoming US president Donald Trump’s government is beginning to take shape, and Earth scientists are getting nervous.

Trump’s latest Cabinet appointments include former Texas governor Rick Perry, a climate sceptic, for energy secretary, and ExxonMobil chief executive Rex Tillerson for secretary of state — a position that would make him the

United States’ lead emissary on climate change. The pair helps to fill out a roster of advisers with strong ties to industry and a distaste for government regulation. Trump’s transition team also asked the Department of Energy (DOE) for the names of employees who had worked on climate-change issues, further unsettling researchers.

“It feels like a war on science, and on climate science in particular,” says Alan Robock, a climatologist at Rutgers University

in New Brunswick, New Jersey. “That’s very upsetting.”

Scientists won a small battle on 14 December, when Trump’s team disavowed the memorandum it sent to the DOE seeking information on climate-change programmes. The request sparked widespread outrage and drew a rebuke from the department after it was leaked on 9 December. At the Fall Meeting of the American Geophysical Union (AGU) last week in San Francisco, California, some researchers billed the episode as a blueprint for how they might defend their interests after Trump takes office on 20 January.

“There is power, even with an administration that never admits a mistake, in bringing things to light,” says Andrew Rosenberg, who heads the Center for Science and Democracy at the Union of Concerned Scientists in Cambridge, Massachusetts.

Other researchers are copying government climate-data sets, to preserve them in case the Trump administration and the Republican-controlled Congress follow through on

proposals to cut back Earth-science research at NASA or otherwise restrict studies of global warming. One rescue effort had archived 11 of 91 data sets on its list for preservation as of 16 December; these include a global temperature record maintained by NASA and palaeoclimate archives held by the National Oceanic and Atmospheric Administration (NOAA).

Marcia McNutt, president of the US National Academy of Sciences, says that private foundations have expressed interest in “funding up to the order of billions of dollars” for climate-change research if the Trump administration reduces support for such work. But McNutt — who directed the US Geological Survey (USGS) from 2009 to 2013 — is not ready to give up on government science. “I don’t want that to be an excuse for the government to pull away — to say private philanthropy can do this, the government doesn’t need to fund it,” she told journalists at the AGU meeting.

The road ahead for scientists looks tough.

Perry dealt with energy issues as governor of Texas, but he lacks experience with key areas of the DOE portfolio, says John Deutch, a chemist at the Massachusetts Institute of Technology in Cambridge. Deutch, who leads the department’s advisory board, says that Trump should identify a deputy energy secretary who understands the agency’s programmes on basic science, nuclear weapons and national security.

And Perry is not the only climate sceptic poised to join Trump’s inner circle. Trump’s pick to lead the US Environmental Protection Agency is Oklahoma attorney-general Scott Pruitt, who has sued the federal government to overturn greenhouse-gas and air-quality rules.

The president-elect has not announced whom he would like to run NASA, NOAA or the USGS, among other science agencies. McNutt says that the National Academies of Science, Engineering, and Medicine have provided his transition team with a list of potential

candidates, but none of those people has been contacted by Trump staff.

Some scientists argue that even if policies to fight climate change are weakened or struck down under Trump, his latest nominations hint that there may be ways to promote clean energy. Tillerson has said that a carbon tax is the best way to address global warming. And although Perry is a strong proponent of fossil fuels, Texas’s wind-power production grew significantly during his governorship.

“Those are places to insert a progressive agenda into an otherwise kind of ugly and cloudy landscape,” says Daniel Kammen, an energy researcher at the University of California, Berkeley.

McNutt advises scientists to stay clear-eyed as they confront whatever challenges the Trump administration brings. “I see so many people in this country freaked out,” she says. “That is exactly what those who want to disrupt science are hoping to achieve.” ■

NAVIGATION SATELLITES

Galileo satellites herald new era for Earth sciences

Europe and Asia will set the atmosphere abuzz with more radio-wave navigation signals.

BY DECLAN BUTLER

After soaring costs and years of delays, Europe’s global satellite-navigation system, Galileo, finally began beaming its first signals to receivers in smartphones and cars on 15 December.

The 18-strong fleet of satellites promises travellers another way to accurately locate their position on Earth, ending Europe’s dependence on the US Global Positioning System (GPS) and Russia’s GLONASS. But Galileo, which was first proposed in 1999, is a big deal for science, too, says Richard Langley, an expert in navigation-satellite systems at the University of New Brunswick in Fredericton, Canada. What most excites scientists is the prospect of combining signals from multiple satellite networks, enabling new kinds of atmospheric and Earth-sciences research.

Galileo’s constellation of satellites should reach its full complement of 30 in 2020, by which time China’s BeiDou system, comprising 35 satellites, is scheduled to enter service. Japan and India are also building regional systems. Altogether, the number of global navigation satellites encircling Earth is set to rise from around 90 today to at least 130 over the

next decade, estimates Oliver Montenbruck, a physicist at the German Aerospace Center in Oberpfaffenhofen, Germany. At the same time, existing satellite fleets will be modernized.

Earth’s atmosphere will then be streaming with many more kinds of radio-wave signal at a greater variety of frequencies — each carrying information about the time and the position of the satellite that sent it. Sat-nav receivers use data from multiple satellites to pinpoint their own position. So simply having more satellites

“The more satellites you have, the greater the precision.”

overhead will help stop signal loss and provide more accurate position fixes, says Langley. “The more satellites you have, the greater the precision,” adds Tonie Van Dam, an Earth scientist at the University of Luxembourg who uses receivers to monitor how Earth’s crust deforms in response to shifting water or ice.

Skies increasingly crowded with radio waves will also benefit weather forecasting and climate research. Scientists use the refraction of navigation-satellite signals in the Earth’s atmosphere to make measurements of atmospheric temperature, pressure, density

and water-vapour content. And the signals can similarly be exploited to measure electron density in the ionosphere, an electrically charged layer in the upper atmosphere. These data are used to track space weather and to monitor tsunamis and earthquakes, says Philippe Lognonné, a geophysicist at the Institute of Earth Physics of Paris. These events disturb air so violently that they send acoustic and gravity waves up to the ionosphere where they perturb electrons. With fully operational Galileo and BeiDou systems, researchers should be better able to estimate tsunami heights, Lognonné says.

Scientists also plan to use multiple navigation-satellite constellations to improve measurements of ocean wind speeds and sea surface roughness, says Jens Wickert, a scientist at the GFZ German Research Centre for Geosciences in Potsdam. Today’s remote-observation ocean maps are built largely by bouncing radar waves off the sea from aircraft or spacecraft, and combining those data with information from other satellite instruments. The best current maps have a spatial resolution of around 80 kilometres and are updated every 10 days. Wickert aims to improve on that using orbiting receivers ►



STEPHANE CORVAJA - ESA

Over the past five years, 18 Galileo satellites have been launched into orbit.

► for navigation-satellite signals. A European experiment called Geros-ISS, which Wickert is leading, aims to fly a receiver on the International Space Station in 2019. The experiment would measure navigation-satellite signals as they reflect off the sea. By combining data from Galileo, BeiDou, GPS and GLONASS, it could map the oceans at

spatial scales down to a few kilometres every four days or less. Many ocean phenomena, such as eddies, occur at these scales, so better maps would help to improve weather and climate-change models.

A fleet of receivers in space could provide even finer resolution. In a step in that direction, on 15 December NASA launched its

own ocean-reflection research mission, the Cyclone Global Navigation Satellite System. A fleet of eight microsatellites, each carrying four navigation-satellite receivers, will measure wind speeds and ocean roughness in the eyes of storms at unprecedented resolutions of a few kilometres every few hours. Chris Ruf, Cyclone's principal investigator and a remote-sensing scientist at the University of Michigan in Ann Arbor, says that the first mission will use GPS only, but he is keen to integrate data from Galileo and BeiDou in follow-ups.

Much research on fusing signals from navigation-satellite systems is taking place under a federation of more than 200 agencies, universities and research centres. Montenbruck, who heads this effort, cautions that it may take more than five years after Galileo and BeiDou enter full service before scientists can exploit their possibilities completely. "Today's use of GPS benefits from 30 years of experience and an excellent understanding and characterization of all the dirty details," he says. "All that still needs to be carried out for Galileo and BeiDou." ■

POLICY

Major rethink for outbreak response

World Health Organization aims to prevent crises similar to the West African Ebola epidemic.

BY ERIKA CHECK HAYDEN

Three years after the start of the world's worst Ebola epidemic, the World Health Organization (WHO) has created a programme to improve its response to disease outbreaks and to prevent another such calamity.

In June, WHO director-general Margaret Chan named medical epidemiologist Peter Salama to lead a new health-emergencies

programme intended to streamline the agency's response to crises. As part of that programme, the WHO has launched the Emerging Diseases Clinical Assessment and Response Network (EDCARN) to provide guidance on how to care for people during disease outbreaks.

Global-health experts say that the changes are a step in the right direction, but both developing and wealthy nations must do much more to avert another devastating epidemic. Some are also concerned that the WHO programme

will have trouble getting the funding it needs to succeed, because of a lack of monetary support from member nations.

"African countries are still so dependent on international and global outfits that the return of Ebola or any other disease will be another déjà vu of national unpreparedness," says virologist Oyewale Tomori at Redeemer's University in Ede, Nigeria.

Tomori says that many developing nations still don't have sufficient capacity for recognizing and responding to an emerging infectious disease.

BUILDING A BRIDGE

The WHO's new programme aims to strengthen local health systems and to bridge global, organizational and governmental efforts to prevent the next outbreak. Daniel Bausch, EDCARN's technical lead, says that the network aims to fill huge gaps exposed during the Ebola crisis: a lack of knowledge about how best to care for people who have such serious diseases, and a shortage of physicians and experts who are prepared to provide that care.


**MORE
ONLINE**

TOP NEWS



Ephemeral antimatter atoms pinned down in milestone laser test go.nature.com/2hphuuj

MORE NEWS

- NASA science chief: 'I have no worries about the resilience of this country' go.nature.com/2hv817k
- Marmosets provide new model for human hearing go.nature.com/2hdkwa2
- Incredibly thin Arctic sea ice shocks researchers go.nature.com/2hp9n45

NATURE PODCAST



Our 2016 round-up: with science carols, holiday reading picks, word games and more nature.com/nature/podcast

CERN

“Clinical management of patients during infectious-disease outbreaks has been one of the neglected areas of public health,” Bausch says. “We’ve realized that it’s one of the many areas where international support is extremely important.”

So EDCARN is bringing together specialists in diseases that are likely to spark outbreaks — including Ebola, Middle East respiratory syndrome (MERS) and Crimean–Congo haemorrhagic fever. These people could be deployed to affected areas along with nurses, logistics experts and infection-control specialists to advise non-governmental agencies, governments and others on the ground who would be caring for patients.

The hope is that this will prevent some of the problems with patient care and health-worker protections seen in the Ebola epidemic, in which many local and foreign health workers were among the 11,310 who died.

FUNDING WORRIES

Public-health policy analysts, such as Lawrence Gostin at Georgetown University in Washington DC, are optimistic about EDCARN. “I think it is a helpful initiative, and is in line with some of the ideas put forward by the Ebola commissions,” says Gostin, who served on two of the five international panels that have recommended major reforms to the



A health worker disinfects a corpse in Monrovia, Liberia, during the recent Ebola outbreak.

global-health system in the wake of the Ebola crisis.

Other health initiatives that were started in the aftermath of the Ebola epidemic have run into problems, which have tempered expectations for EDCARN. A US-led effort to boost domestic health systems has met with local resistance, and a WHO-led programme with

similar goals does not have enough long-term funding to keep going, Gostin says.

To succeed, EDCARN must raise US\$485 million for the 2016–17 fiscal year, and is currently only 56% funded. Gostin hopes that stinginess among the WHO member countries won’t doom the fledgling attempts to head off another crisis. ■

366 DAYS:
the year in science



An emperor penguin in Antarctica's Ross Sea, which will now host the world's largest marine reserve.

2 0 1 6

I N R E V I E W

FROM AN ELECTION THAT STUNNED THE WORLD TO CATASTROPHIC TECHNICAL GLITCHES IN SPACE, RESEARCHERS WEATHERED A TURBULENT YEAR. BUT THEY ALSO ANNOUNCED SOME REMARKABLE ADVANCES — THE DIRECT DETECTION OF GRAVITATIONAL WAVES, THE BIRTH OF A BABY WITH DNA FROM THREE PEOPLE AND AN ARTIFICIAL INTELLIGENCE THAT CRACKED THE ONE BOARD GAME THAT COMPUTERS HAD YET TO MASTER.

PAUL NICKLEN/NGC

CATCHING A WAVE

Physicists bagged some big game this year. On 11 February, researchers announced that they had finally sensed the ripples in the structure of space-time known as gravitational waves — capping a decades-long quest. The signal, spotted in September 2015 by the twin detectors of the Laser Interferometer Gravitational-Wave Observatory (LIGO) in Louisiana and Washington state, came from the merger of two black holes some 1 billion years before.

The announcement was a stunning affirmation of Albert Einstein's general theory of relativity, almost 100 years after he had published it. And it provided the most direct evidence yet that black holes — another prediction of Einstein's theory — exist. Astrophysicists hailed LIGO's feat as a triumph, saying that it heralded a new way of observing the cosmos, enabling the detection of phenomena that might not be picked up by other means.

Just weeks after the LIGO announcement, another experiment demonstrated that the search for gravitational waves could one day occur in space. The European Space Agency's LISA Pathfinder mission tested technologies for a future trio of probes that would sense gravitational waves coming from even larger and more-distant objects than the ones LIGO observed.

Particle physicists were not so lucky. They spent much of the year holding their collective breath. Two separate experiments at the Large Hadron Collider near Geneva, Switzerland, had reported anomalous measurements in late 2015 that suggested the existence of a particle six times as massive as the Higgs boson.

At the time, experimenters warned that the anomalies could be statistical flukes. And more data released in August confirmed this. By then, theoreticians had written hundreds of papers in attempts to interpret the original data with a zoo of possible models.

NEW WORLD ORDER

A tumultuous US presidential campaign ended in a surprise victory for Republican businessman Donald Trump in November. Researchers struggled to understand how a Trump administration would treat science — in part, because it did not feature prominently on the campaign trail. Still, some of Trump's views were clear: he has alleged that climate change is a hoax perpetrated by the Chinese, and has pledged to withdraw the United States from the Paris climate agreement. He has also suggested a link between autism and childhood vaccinations.

As Trump's administration began to take shape, researchers started to lobby against what they saw as an incoming president with little

President-elect Donald Trump on the campaign trail in October.

use for science. In late November, more than 2,300 scientists — including 22 Nobel laureates — sent a letter to Trump urging him to “adhere to high standards of scientific integrity and independence in responding to current and emerging public health and environmental health threats”.

The United Kingdom's 23 June vote to leave the European Union shook the country's scientific community. Researchers remain worried about the fate of millions of euros in annual research funding from the EU and the immigration status of UK university staff from non-British EU nations. Some UK researchers have reported being cut out of EU collaborations, and some foreign scientists say that they feel so unwelcome that they plan to leave the country. In happier news, the UK government — led by a cabinet that came to power in the wake of the Brexit vote — announced a surprise funding boost in November worth £2 billion (US\$2.5 billion) annually by 2020.

A failed military coup in Turkey in July spelt upheaval of a different kind for academics: the Council of Higher Education promptly fired more than 1,500 university deans. About 58% of the positions have since been refilled by their former occupants. But more than 6,500 professors have been dismissed on suspicion of involvement in the coup. Human-rights groups say that many of the dismissed are innocent.

Political and economic woes rocked Venezuela, Brazil and South Africa this year — and did not spare researchers. Rolling blackouts, food queues and increasing violence prompted hundreds of scientists to leave Venezuela's universities, and, in some cases, the country.

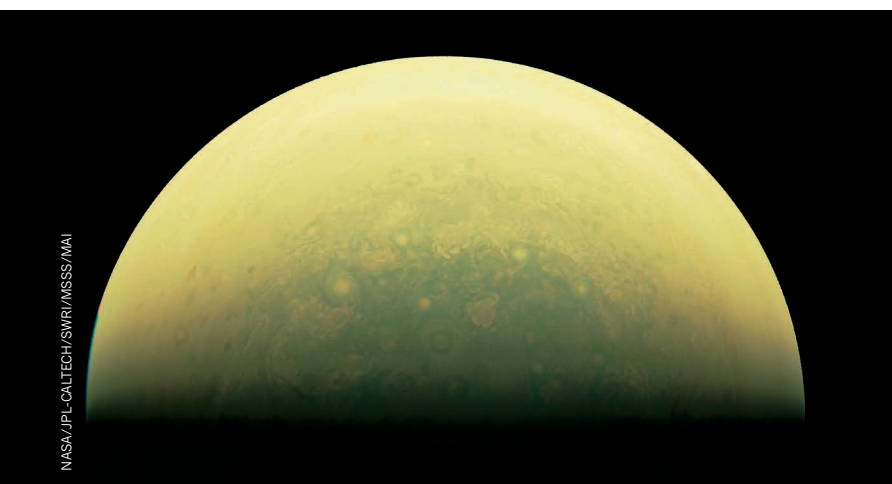
Brazil's researchers are facing drastic budget cuts and the demotion of the science ministry, and protested against proposals to freeze federal science spending and weaken the country's environmental laws. And austerity measures in South Africa have led to chronic underfunding of universities and triggered a rash of campus protests and violence.

TO BOLDLY GO

In the year that saw the 50th anniversary of *Star Trek*, technical glitches set back several space missions — but there were also notable victories. In March, the Japan Aerospace Exploration Agency's flagship Hitomi X-ray astronomy satellite failed just weeks after launch. Investigators determined that a software error had caused the spacecraft to rotate out of control and break apart. In July, NASA's Juno probe arrived at Jupiter, but problems with its main engine delayed the rocket firing ▶



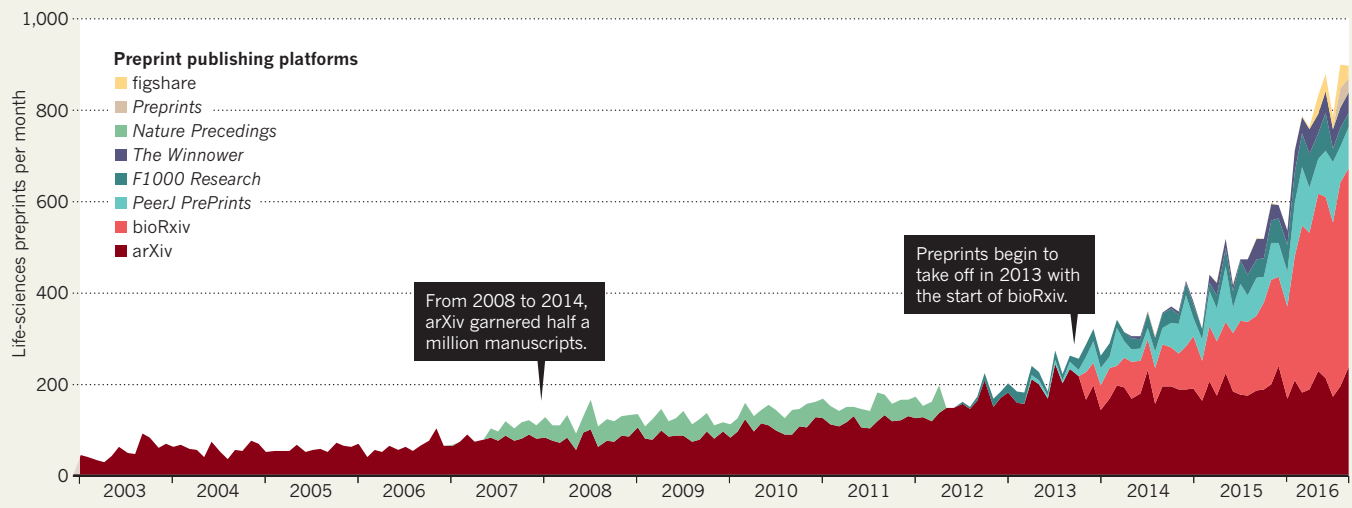
WIN MCNAMEE/GETTY



A sunlit part of Jupiter.

BIOLOGY PREPRINTS ON THE RISE

Life scientists have not embraced preprints — studies published online before peer review — as readily as have physicists or mathematicians. But that is changing. Even the US National Institutes of Health is getting behind the practice, mandating that grant recipients in its 4D Nucleome consortium publish preprints of their work.



SOURCE: ASAPBIO/PREPRINTED/BIOXIV

► that would have shrunk its orbit into a tighter ellipse around the planet. The spacecraft continues to gather data on Jupiter's atmosphere and magnetosphere on every fly-by — just more slowly than planned.

Meanwhile, the European Space Agency (ESA) mourned the end of two probes. In October, the Schiaparelli lander — part of the ESA's ExoMars mission — smashed onto the red planet's surface after a measurement error caused its parachute and braking rockets to deploy at the wrong times. But at least its companion spacecraft managed to enter orbit around Mars. ESA's other loss was sad for some scientists, but deliberate. The pioneering Rosetta spacecraft crashed onto the surface of Comet 67P/Churyumov–Gerasimenko in September as planned, radioing back close-up images before it lost contact — and bringing the mission to an end.

Rising space power China garnered several wins. In August, it launched the first ever quantum satellite, aimed at testing ways to extend secure quantum communication into space. In September, the country completed construction on the world's largest single-dish telescope, the Five-hundred-meter Aperture Spherical Radio Telescope in the southwestern province of Guizhou. And in November, China launched the Long March 5 rocket, one of the world's most powerful. It is meant to send people, rovers and heavy-duty planetary probes into space.

Finally, two Chinese astronauts broke their country's record for the longest-duration space mission when they spent a month aboard the Tiangong 2 space laboratory in October and November.

CRISPR IN COURT

The development of new applications for the genome-editing tool CRISPR–Cas9 continued apace. On 28 October, a patient with lung cancer at West China Hospital in Chengdu became the first person to be treated with cells edited using CRISPR–Cas9. As part of a clinical trial, researchers disabled a gene that normally holds a cell's immune system in check, in the hope that the edited cells would mount an immune response against the cancer. More cancer trials using treatments based on CRISPR–Cas9 are expected in the United States and China next year.

But the commercial landscape for CRISPR–Cas9 therapies remains uncertain. The battle over US patent rights to the gene-editing technique reached fever pitch after the US Patent and Trademark Office declared an 'interference' proceeding between two research teams in January. The proceeding, which could conclude early next year, aims to determine

who first invented the technique. Each team applied for patents that could be crucial for commercial applications.

Meanwhile, research using CRISPR–Cas9 in human embryos expanded this year. It is a controversial area of research that has raised concerns about the potential for designer babies — but regulators in some countries have approved projects in this field. Teams in China, the United Kingdom and Sweden announced their intentions to use the technique to optimize its use in embryos and to study human development. Work in the United States is expected to follow, despite a prohibition on the use of federal funds to study human embryos or to modify human eggs or sperm.

CLIMATE CRUNCH

Representatives of a record 174 countries and the European Union gathered on Earth Day, 22 April, to sign the international climate agreement forged in Paris in December 2015. But for the accord to come into force, more than 55 countries accounting for at least 55% of global greenhouse-gas emissions needed to submit ratification or acceptance documents.

The biggest boost came in September, when the United States and China — which together account for 38% of global emissions — formally joined the agreement. Brazil and 30 other countries joined a few weeks later, and the EU sealed the deal on 5 October. The pact came into effect on 4 November.

But that wasn't the only global climate deal afoot. On 6 October, the United Nations' International Civil Aviation Organization curbed emissions from international flights. And on 15 October, 197 countries agreed to amend the Montreal Protocol — designed to protect the ozone layer — to phase out hydrofluorocarbons, powerful greenhouse gases commonly used in air conditioners. Countries also broke a four-year-long impasse on

Green lights illuminate the Eiffel Tower to celebrate the Paris climate agreement.

GEOFFROY VAN DER HASSELT/ANADOLU AGENCY/GETTY

GOOGLE/GETTY

28 October to create the world's largest marine reserve in the Ross Sea off the coast of Antarctica.

All the while, global warming continued. An epic El Niño in the tropical Pacific Ocean helped set global-temperature records in the first five months of the year. This put 2016 on track to become the third straight warmest year in a row. The blazing warmth prompted corals around the world to bleach, a process in which the stressed animals expel the algae that help to keep them alive. The El Niño faded away in May, but coral bleaching continued throughout the year and is expected to continue into 2017.

ZIKA SPREADS

In February, the World Health Organization (WHO) declared that clusters of birth defects linked to outbreaks of Zika virus in Brazil constituted a global public-health emergency. These birth defects included severe cases of microcephaly, a condition in which fetuses or newborns have abnormally small heads and brains.

But the expected explosion in microcephaly cases and other Zika-linked birth defects across the Americas has not materialized, despite the virus's spread across the continents.

Even in Brazil, extremely high rates of microcephaly remained confined to the country's northeast region — and researchers began to suspect the influence of confounding factors. In July, Brazilian authorities launched a study to find out whether environmental, socio-economic or biological elements, when combined with Zika infection, could explain the odd distribution of elevated rates. They expect preliminary results early next year.

On 18 November, the WHO declared the end of the international public-health emergency on the grounds that the link between Zika and birth defects had been established, and that the focus needed to shift to understanding the consequences of Zika infections, including the birth defects, and developing a vaccine.

Several ongoing international research projects should produce results next year on a number of Zika-related questions, such as what proportion of infected pregnant women go on to have babies with birth defects.



Go player Lee Sedol reviews a match against artificial intelligence.

MIND GAMES

In January, a computer program beat a world-class human player at the ancient game of Go for the first time. But the ultimate showdown was in March, when the artificial intelligence (AI), called AlphaGo, trounced Lee Sedol — one of the world's top players. The AI, developed by the Google-owned company DeepMind in London, opened with three consecutive wins in the five-round tournament. Lee took the fourth game and nearly won game five, but AlphaGo triumphed.

In October, DeepMind researchers debuted another AI, one capable of navigating the London Underground without any previous knowledge. The sophisticated program combined memory with the ability to learn from experience. This brought AI a step closer to performing human-like tasks such as reasoning.

AI also helped to reduce errors in machine translation of languages by around 60%, and aided physicists looking for new super materials. These advances were largely powered by deep learning, which harnesses huge data sets and a hierarchical, brain-like method of computing.

CONTROVERSIAL CONCEPTION

After decades of research, assisted-reproduction techniques that mix DNA from three people are bearing fruit. These procedures prevent children from inheriting metabolic diseases caused by flaws in mitochondria, the cell's energy-producing structures.

In September, researchers working in a Mexican clinic reported the birth of the first healthy baby conceived through one such procedure. A baby in China was also reportedly born using the same technique. And in October, a clinic in Ukraine announced that two previously infertile women had conceived through a similar procedure. On 15 December, following scientists' advice, the United Kingdom's Human Fertilisation and Embryology Authority said that the technique was ready for clinical use, which could start as soon as 2017.

Written by Alison Abbott, Declan Butler, Davide Castelvecchi, Daniel Cressey, Elizabeth Gibney, Heidi Ledford, Jane J. Lee, Lauren Morello, Sara Reardon, Jeff Tollefson and Alexandra Witze.

NACHO DOCE/REUTERS



A four-month-old baby born in Brazil with microcephaly.

366 DAYS:
the year in science

IMAGES OF THE YEAR

In a year of political turmoil and shock, science, too, came up with surprises. To document some of these wonders, photographers roamed the world, revealing objects from the microscopic to the cosmic in scale.

Images selected by *Nature's* art and design team
Text by Daniel Cressey

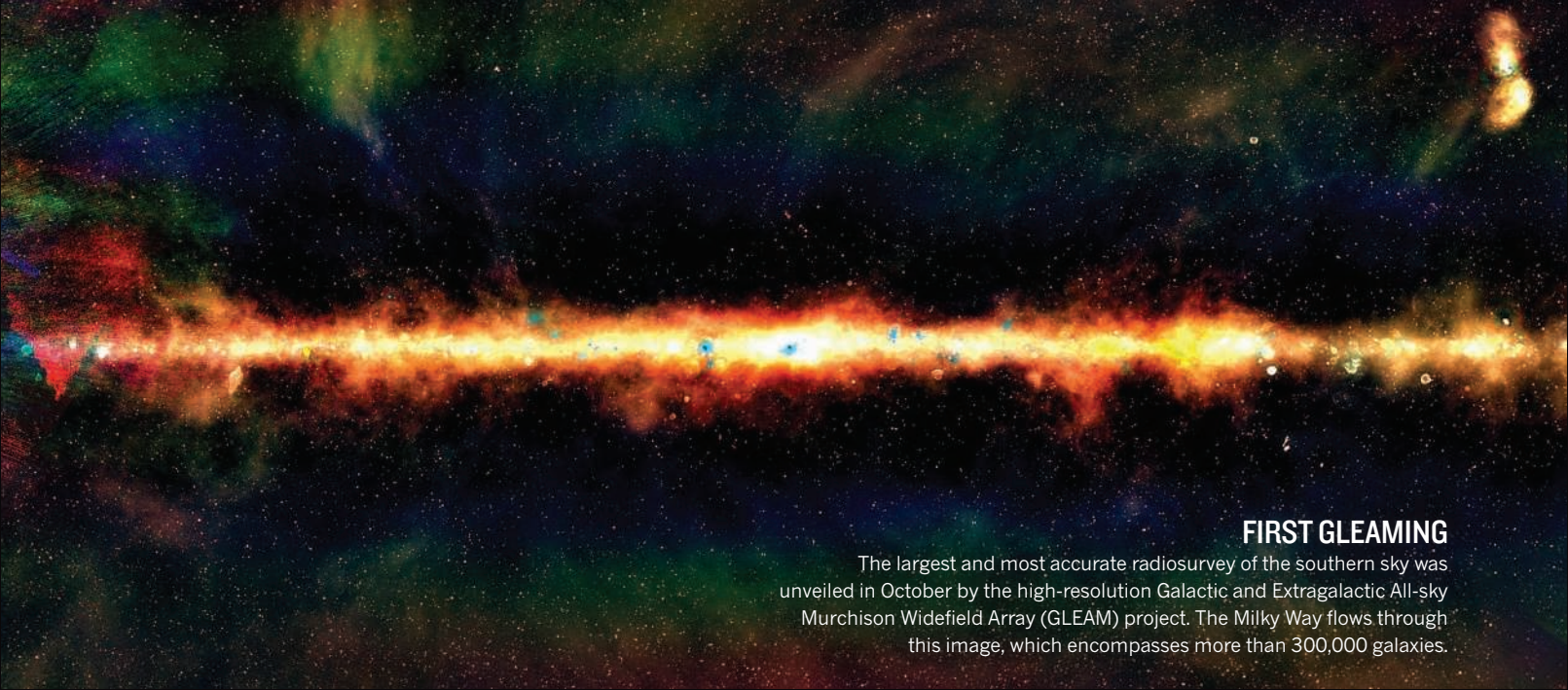
STRIKING CRANES

Hundreds of thousands of sandhill cranes (*Grus canadensis*) converge on Platte River in Nebraska as part of their annual migration. Photographer Randy Olson was taking long-exposure shots in March when lightning struck, creating these ghostly outlines.



ANCIENT IVORY

The vast tusk of a long-dead mammoth is carried out of a forest in Yakutia, Siberia. Ancient ivory from mammoths has become so valuable that some prospectors now illegally 'mine' them from permafrost. A large tusk can be worth tens of thousands of dollars.



FIRST GLEAMING

The largest and most accurate radiosurvey of the southern sky was unveiled in October by the high-resolution Galactic and Extragalactic All-sky Murchison Widefield Array (GLEAM) project. The Milky Way flows through this image, which encompasses more than 300,000 galaxies.



EN ROUTE

This long-exposure shot shows the November launch of a Soyuz spacecraft from the Baikonur Cosmodrome in Kazakhstan. It ferried Peggy Whitson, Oleg Novitskiy and Thomas Pesquet to the International Space Station.



FANTASTIC FOOT

This spectacular tarsus — the lowermost segment of an insect leg — is roughly 2 millimetres in diameter and belongs to a male diving beetle, which uses it to attach to a female's back during mating.



CHINA CHANGES

China this year revealed ambitious plans to cut coal use and pollution and to embrace renewable energy. But this steel plant in Inner Mongolia is just one example of the many industries that stand in the way of that reform.

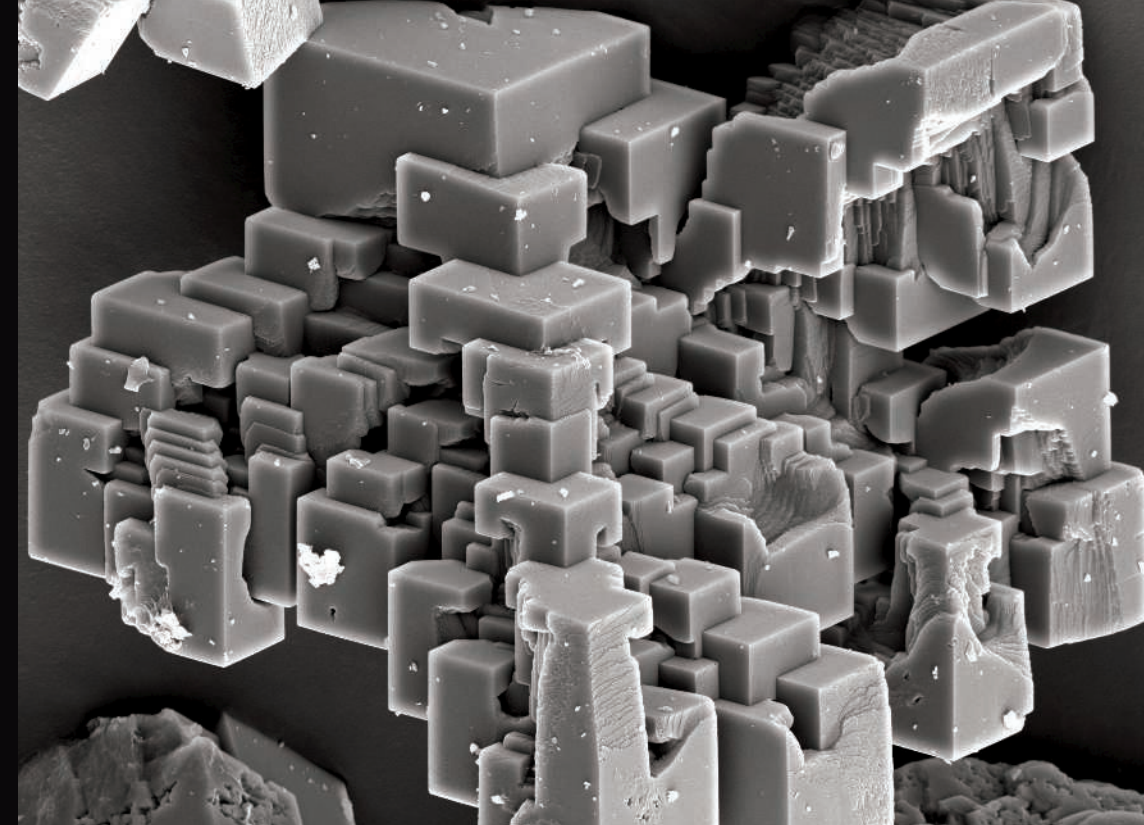
SPACE STORMS

Far below the International Space Station, lightning flashes illuminate the clouds, as human activity is revealed by clusters of lights. Two Russian spacecraft visiting the station can be seen in the foreground.

CRANES: RANDY OLSON/NATIONAL GEOGRAPHIC; TUSK: AMOS CHAPPEL/RADIO FREE EUROPE; SOYUZ: ZUMA PRESS/EVJIVE; BEETLE: IGOR SIWANOWICZ/NIKON SMALL WORLD 2016; GLEAM: NATASHA HURLEY-WALKER (CURTIN/CORBIS); GLEAM TEAM: CHINA: KEVIN FRAYER/GETTY; ISS: NASA; CRYSTAL: CHRISTINE KIMPTON/RPS INTERNATIONAL IMAGES FOR SCIENCE 2016; MUMMY: ANNE AUSTIN; DISCO: C. PAN ET AL./NATURE METHODS; STEM CELL: S. A. FERREIRA, C. LOPO & E. GENTILEMAN, KCL/WELLCOME

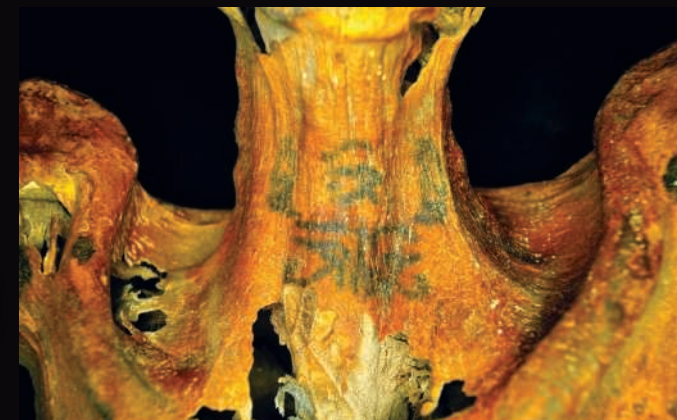
CRYSTAL STEPS

These strange structures are calcium carbonate crystals, imaged at 2,000× magnification.



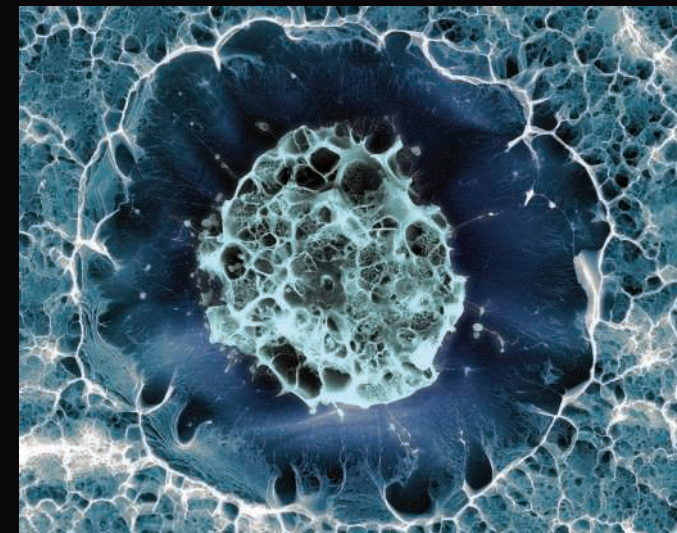
SACRED SYMBOLS

In April, remarkable images of ancient Egyptian tattoos found on a mummy were shown at a meeting of the American Association of Physical Anthropologists. The tattoos include two seated baboons and a symbol of protection on the mummy's neck.



SEE-THROUGH AND SMALL

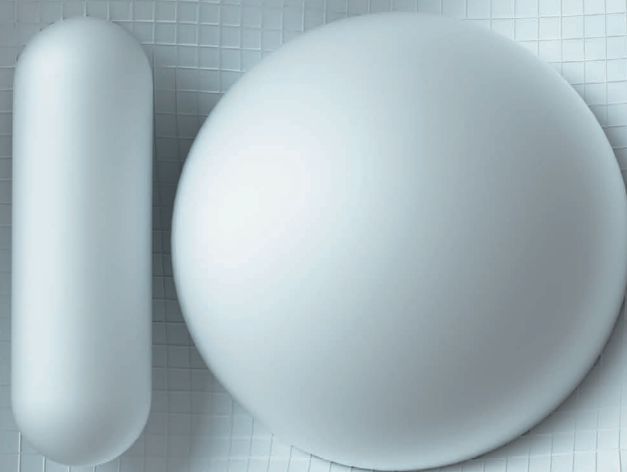
In August, a team in Germany unveiled 'ultimate DISCO' — a technique that both renders tissues transparent and shrinks specimens, so that a whole animal can be imaged in one go. The technique can reveal the nervous system and organ systems within a body in unprecedented detail.



STRIKING CELL

This human stem cell is just 15 micrometres across, and was false-coloured after being imaged using cryogenic scanning electron microscopy.

TEN PEOPLE WHO MATTERED THIS YEAR



NATURE'S 10

GABRIELA GONZALEZ / DEMIS HASSABIS / TERRY HUGHES

GUUS VELDER / CELINA M. TURCHI / ALEXANDRA ELBAKYAN

JOHN ZHANG / KEVIN ESVELT / GUILLEM ANGLADA-ESCODÉ / ELENA LONG

366 DAYS:
the year in science



GABRIELA
GONZALEZ

GRAVITY SPY

A physicist helped to catch the first direct signs of long-sought gravitational waves.

BY DAVIDE CASTELVECCHI

A year ago, Gabriela Gonzalez was struggling to contain the biggest secret of her life. Two giant detectors in the United States had picked up signs of gravitational waves — wrinkles in space-time imagined by Albert Einstein but never before directly witnessed. It was Gonzalez's job to help lead more than 1,000 scientists in their careful efforts to verify the discovery before announcing it to the public.

News like that doesn't stay under wraps for long, but the discovery was so momentous that the research team took nearly five months to analyse data from the two Laser Interferometer Gravitational-Wave Observatory (LIGO) detectors in Washington state and Louisiana. As spokesperson for the LIGO Scientific Collaboration, Gonzalez was one of the key people coordinating the analysis by groups scattered around the world, including researchers at the Virgo interferometer near Pisa, Italy, which pools its data with LIGO.

The role of shepherding this massive effort made use of Gonzalez's multidimensional talents. Most physicists know early on whether they will be a theorist or an experimentalist. But Gonzalez started her graduate studies as a theoretical physicist and only later switched to experimental work, when she showed uncommon aptitude. "It was the thing that set her up as a first-class scientist," says Rainer Weiss, a physicist at the Massachusetts Institute of Technology in Cambridge and one of the founders of LIGO.

Throughout her career, Gonzalez has done "a bit of everything" at LIGO, she says. For a while, she took on the crucial task of diagnosing

the performance of the interferometers to make sure that they achieved unparalleled sensitivity — which is now enough to detect length changes in the 4-kilometre-long arms of the interferometers to within one part in 10^{21} , roughly equivalent to the width of DNA compared with the orbit of Saturn. She has helped to lead the teams that analyse the data. And she nudged gravitational-wave researchers and dozens of their colleagues in conventional astronomy into signing pacts of cooperation. Together, they will look for phenomena that emit both gravitational and electromagnetic waves, in what has been called the coming age of multimessenger astronomy.

WILLIAM WIDMER/REDUX/EYEVINE

In the hectic months before announcing the LIGO discovery, Gonzalez and her colleagues struggled to make sure that they had iron-clad evidence. They knew that history had not been kind to those who had previously reported gravitational waves. Most recently, in early 2015, an international collaboration had to retract its claims that a telescope at the South Pole had discovered indirect signs of the long-sought vibrations.

To add to the pressure on the LIGO team, rumours of a discovery began to leak within a week of the initial finding, and reporters started to call. Throughout the long analysis period, Gonzalez says, she never made an important decision without consulting colleagues. But others laud her leadership. "What Gaby did is, she managed to get us through this period," Weiss says.

Gonzalez is based at Louisiana State University in Baton Rouge, close to the LIGO interferometer in Livingston. In 2008, she became the first woman to receive a full professorship in her department. She says that she has never experienced outright sexual harassment or discrimination during her career, but "I had to prove myself perhaps more than other people".

Gonzalez has said that after her current term as LIGO spokesperson ends in March 2017, she will not run again. She plans to go back to full-time research. The field of science she helped to create — gravitational-wave astronomy — has just seen its dawn. "It has always been a fun ride. And now it's even better." ■

MIND CRAFTER

An AI developer beat one of the best at Go. Next up, solve global problems.

BY ELIZABETH GIBNEY

For veteran gamer Demis Hassabis, March brought the toughest match of his life — and he wasn't even playing. Hassabis had to watch from the sidelines as his team's creation, the computer program AlphaGo, took on Lee Sedol, a top-ranked champion in the strategy game Go. The computer won, marking a huge victory for the field of artificial intelligence (AI) and another in a series of triumphs for Hassabis.

As co-founder of DeepMind, the London-based firm that developed AlphaGo, Hassabis was both elated and relieved. "It felt like our moonshot, and it was successful," he says.

But the win was about much more than Go. Hassabis wanted to show the world the power of machine-learning techniques, which he hopes to someday harness in a human-like, general AI capable of solving complex global problems.

Hassabis had sketched this vision out as a precocious youth. A chess prodigy, he began designing innovative, multimillion-selling video games while in his teens and

started his own company in his early 20s. After completing a PhD in cognitive neuroscience, he founded DeepMind in 2010. Google bought the company 4 years later for a reported £400 million (more than US\$650 million at the time).

At the firm, researchers apply inspiration from neuroscience to eye-catching AI tasks, from synthesizing speech to navigating the London Underground. Each algorithm builds complexity on to the last, says Hassabis, and weaves in capabilities that have historically been developed separately in AI. DeepMind AIs have gone from learning how to see, and acting on that vision, to using it to plan and reason. In terms of real-world problem-solving, the team used machine learning to cut power usage in Google's data centres by 15%, something that Hassabis hopes to apply on a much grander scale.

Although the company's researchers do publish, their work-in-progress is kept under wraps, which irks some academics. And some data-privacy advocates have concerns over Google DeepMind's plans to collaborate with the UK National Health Service. Scientists, however, have been flocking to work at the company.

In person, Hassabis is unassuming but eager. He has a knack for swaying others to his passion, says Eleanor Maguire, his former PhD supervisor at University College London. "Once he gets talking about something he's interested in, it's infectious," she says. Fitting research alongside running the company now means saving science for the small hours of the morning, something Hassabis says he doesn't mind. "It's a very important mission that we're on, and I think it's worth the sacrifice." ■



DEMIS
HASSABIS



REEF SENTINEL

A coral researcher sounded the alarm over massive bleaching at the Great Barrier Reef.

BY DANIEL CRESSEY

When Terry Hughes flew over the Great Barrier Reef in March, his heart sank at the sight of telltale pale patches just below the surface, where corals were dead or dying.

Hughes, director of the Australian Research Council's (ARC's) Centre of Excellence for Coral Reef Studies in Townsville, says that he and his students wept after looking at the aerial surveys of the damage. The bleaching hit nearly all of the reef, with initial surveys showing 81% of the northern section suffering severely. It was the most devastating bleaching ever documented on the Great Barrier Reef — and part of a wider event that was harming corals across the Pacific.

The trigger for this year's coral troubles in the Pacific was a strong El Niño warming pattern in the tropical part of that ocean. Abnormally high water temperatures prompt corals to expel the symbiotic zooxanthellae algae that provide them with much of their food — and their colour. Some corals can recover after bleaching, but others die. Follow-up studies in October and November found that 67% of shallow-water corals in the 700-kilometre northern section of the Great Barrier Reef had died.

When the massive El Niño reared up in the Pacific in 2015, Australian researchers feared that the country's reefs could be in danger. So Hughes, one of the world's leading coral researchers, assembled a task force ready

ANDREW RANKIN FOR NATURE

COOLING AGENT

An atmospheric chemist laid the foundation for an international climate agreement.

BY JEFF TOLLEFSON

It isn't often that atmospheric chemists get to help save the world, but Guus Velders had his chance in October. He was attending international negotiations in Kigali, Rwanda, that were seeking to phase out production and use of hydrofluorocarbons (HFCs), extremely potent greenhouse gases commonly used in air conditioners.

Most nations had agreed on an aggressive timetable to begin eliminating the compounds, but India and a handful of other countries wanted an extra four years. After plugging the numbers into a model on his laptop computer, Velders informed negotiators that this particular concession would have little impact on the planet.

That and his earlier work helped to smooth the way for a widely hailed global accord, which was signed on 15 October. Velders, a soft-spoken researcher at the National Institute for Public Health and the

Environment in Bilthoven, the Netherlands, is proud of the part he played. "I've never been involved in a process that leads to a global agreement on climate before," he says.

It was no coincidence, however. Colleagues say that Velders has become the world's expert on HFC emissions, and that nobody else could have provided such rapid analysis in Kigali. He is part of a community of scientists that has helped to refashion the 1987 Montreal Protocol — an international agreement designed to protect the stratospheric ozone layer — into a tool with which to fight global warming.

The refrigerants that fall within the scope of the protocol are also powerful greenhouse gases, and Velders' team showed that the Montreal agreement actually did more to control global temperatures than did the 1997 Kyoto Protocol climate treaty. More recently, the team projected how much warming HFCs were likely to cause over the twenty-first century. That helped to set the stage for the agreement on HFCs, which was reached as an amendment to the Montreal Protocol.

"The Velders team always answered the right questions at the right time," says Durwood Zaelke, president of the Institute for Governance & Sustainable Development, an advocacy group in Washington DC. "It's safe to say that we wouldn't have this agreement without them."

Now it's back to the drawing board for Velders' team. Their scenario about how HFC emissions would grow over time was rendered obsolete by the new agreement to ban them. That's the kind of intellectual setback that Velders heartily accepts. ■

to survey the reef if bleaching occurred. The group eventually expanded to 300 scientists. “We put together a very detailed research plan, hoping of course that it wouldn’t happen,” he says.

Hughes is based close to the central portion of the Great Barrier Reef. After leading the initial surveys, he became the de facto spokesperson on the catastrophe. At the height of media interest in the bleaching, Hughes did 35 interviews in one day.

“In Australia, even people who have never been to the Great Barrier Reef and might never go there regard it as an icon,” says Bob Pressey, a fellow researcher at the ARC centre.

The crisis on the reef defied some rules. Conventional thinking on bleaching events, says Hughes, is that corals die slowly from starvation after their zooxanthellae leave. But this year, water temperatures were so high that “we saw a lot of corals die before the starvation kicked in. They actually cooked.”

Corals throughout the world have struggled in the past couple of years, as global temperatures have repeatedly hit record highs. In October 2015, the US National Oceanic and Atmospheric Administration declared that a global bleaching event was happening as coral reefs in Hawaii, Papua New Guinea and the Maldives began to succumb.

This year, the bleaching spread to Australia, Japan and other parts of the Pacific. Researchers say that, as climate change drives up baseline temperatures, bleaching will afflict reefs more frequently. Under some scenarios, this could happen so often that most corals can no longer survive.

Hughes is not ready to give up on the Great Barrier Reef just yet. But the recent bleaching has left corals in a weakened state, prone to attacks from pathogens and predators. Another bleaching event in the near future could bring further damage. “The message to people,” he says, “should be we’ve got a closing window of opportunity to deal with climate change.” ■



ZIKA DETECTIVE

A physician raced to make sense of a medical mystery in northeast Brazil.

BY DECLAN BUTLER

Fears about the Zika virus spread across the globe in 2016, and the epicentre of concern was Brazil, where the epidemic first appeared in the Americas. Some researchers even called for postponing the Olympic Games scheduled for Rio de Janeiro in August that year. But away from the media frenzy, Celina Maria Turchi Martelli battled on the front lines in northeast Brazil to make sense of the medical mystery there.

Turchi, a physician and infectious-disease expert, has had her life turned upside down by Zika since September 2015. That’s when the ministry of health asked her to investigate a sharp rise in reports of babies born with abnormally small heads and brains, a condition known as microcephaly, in her home state of Pernambuco. She quickly became convinced that the country was facing a public-health emergency. “Not even in my worst nightmare as an epidemiologist had I imagined a microcephaly neonate epidemic,” she says.

Turchi, who is based at the Aggeu Magalhães Research Center in Recife, immediately contacted scientists across the globe for help. She formed a networked task force of epidemiologists, infectious-diseases experts, paediatricians, neurologists and reproductive biologists. The challenges were formidable, says Turchi: there were no reliable lab tests for Zika, and there was no consensus on a case definition of microcephaly. But the intense networking paid off, and Turchi and her colleagues eventually generated enough evidence to demonstrate a link between the condition and infection with Zika in the first trimester of pregnancy.

Still, the mysteries are far from solved, says Turchi. Although Zika has spread across the Americas, the expected explosion in the number of microcephaly cases outside northeast Brazil has not materialized. Turchi and her task force are now trying to work out why. When she started going into the hospitals of Recife to investigate the outbreak, Turchi says, she had to innovate. “There was no book to follow.” Now, she and her colleagues are writing that book. ■



PAPER PIRATE

The founder of an illegal hub for paywalled papers has attracted litigation and acclaim.

BY RICHARD VAN NOORDEN

It took Alexandra Elbakyan just a few years to go from information-technology student to famous fugitive.

In 2009, when she was a graduate student working on her final-year research project in Almaty, Kazakhstan, Elbakyan became frustrated at being unable to read many scholarly papers because she couldn't afford them. So she learnt how to circumvent publishers' paywalls.

Her skills were soon in demand. Elbakyan saw scientists on web forums asking for papers they couldn't access — and she was happy to oblige. "I got thanked many times for sending paywalled papers," she says. In 2011, she decided to automate the process and founded Sci-Hub, a pirate website that grabs copies of research papers from behind paywalls and serves them up to anyone who asks. This year, interest in Sci-Hub exploded as mainstream media cottoned on to it and usage soared. According to Elbakyan's figures, the site now hosts around 60 million papers and is likely to serve up more than 75 million downloads in 2016 — up from 42 million last year and, by one estimate, encompassing around 3% of all downloads from science publishers worldwide.

It is copyright-breaking on a grand scale — and has brought Elbakyan praise, criticism and a lawsuit. Few people support the fact that she acted illegally, but many see Sci-Hub as advancing the cause of the open-access movement, which holds that papers should be made (legally) free to read and reuse. "What she did is nothing short of awesome," says Michael Eisen,

AP/NET JOLLY/FICKR/CC BY 2.0

FERTILITY REBEL

A physician jump-started debate over a controversial IVF procedure.

BY SARA REARDON

Shock, anger, scepticism and congratulations. Those were some of the reactions that fertility specialist John Zhang triggered in the scientific community in September, when he announced that a controversial technique that mixes DNA from three people had been used to produce a healthy baby boy.

This kind of technique is intended to prevent children from inheriting disorders involving mitochondria — the cellular structures that produce energy. But ethical and safety concerns have prompted the United States to ban such procedures without a permit. Zhang, who works at New Hope Fertility Center in New York City, performed the technique at the company's clinic in Mexico.

Critics saw this as an attempt to evade regulation, and complained that he had announced the work at a conference rather than in a publication.

But Zhang brushes aside those objections. "The most important is to have a live-birth baby, not to tell the whole world," he says.

Zhang has a habit of pushing scientific and ethical boundaries. In the 1990s, he worked with reproductive endocrinologist Jamie Grifo at the New York University Langone Medical Center to develop a version of the technique that Zhang used this year. The approach was designed to help older women to become pregnant by replacing their ageing mitochondria with those from younger eggs. No successful pregnancies resulted.

When US regulators began restricting this technique in 2001, Zhang and his collaborators in China took over the work. In 2003, Zhang's team created and implanted multiple embryos into a woman. After all the fetuses were miscarried, China banned the technique as well.

Grifo and some others applaud Zhang's latest work. "I think it's a great thing it was finally done," says Grifo. But others have criticized the New Hope team. "A lot of things they did were completely unsafe," such as infusing the donor's egg with a drug that could cause chromosomal abnormalities, says Shoukhrat Mitalipov, a stem-cell scientist at Oregon Health & Science University in Portland.

Zhang is undeterred. He says that plenty of other families at risk of mitochondrial disease have expressed interest in his procedure, and he hopes to perform it in other countries. "Five to ten years from today, people will look at it and say, 'Why were we all so stupid, why were we against it?'" he says. "I think you have to show the benefit to mankind." ■

a biologist and open-access supporter at the University of California, Berkeley. “Lack of access to the scientific literature is a massive injustice, and she fixed it with one fell swoop.”

For the first few years of its existence, the site flew under the radar — but eventually it grew too big for subscription publishers to ignore. In 2015, the Dutch company Elsevier, supported by the wider publishing industry, brought a US lawsuit against Elbakyan on the basis of copyright infringement and hacking. If Elbakyan loses, she risks having to pay many millions of dollars in damages, and potentially spending time in jail. (For that reason, Elbakyan does not disclose her current location and she was interviewed for this article by encrypted e-mail and messaging.) In 2015, a US judge ordered Sci-Hub to be shut down, but the site popped up on other domains.

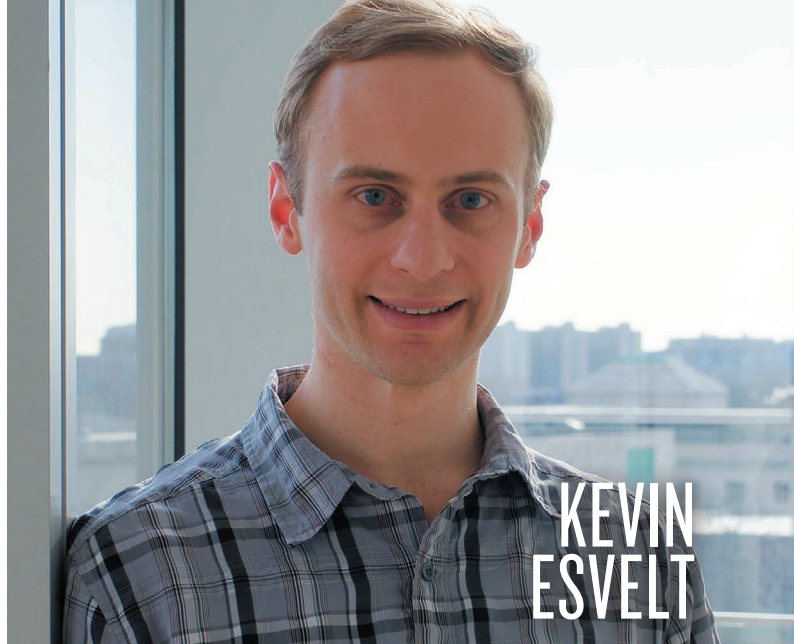
Elbakyan has found her name splashed across newspapers, and says she typically gets a hundred supportive messages a week, some with financial donations. She says she feels a moral responsibility to keep her website afloat because of the users who need it to continue their work. “Is there anything wrong or shameful in running a research-access website such as Sci-Hub? I think no, therefore I can be open about my activities,” she says.

Critics and supporters alike think that the site will have a lasting impact, even if it does not last. “The future is universal open access,” says Heather Piwowar, a co-founder of Impactstory, a non-profit firm incorporated in Carrboro, North Carolina, which helps scientists track the impact of their online output. “But we suspect and hope that Sci-Hub is currently filling toll-access publishers with roaring, existential panic. Because in many cases that’s the only thing that’s going to make them actually do the right thing and move to open-access models.”

Whether or not that’s true, Elbakyan says she will keep building Sci-Hub — in particular, to expand its corpus of older manuscripts — while studying for a master’s degree in the history of science. “I maintain the website myself, but if I’m prevented, somebody else can take over the job,” she says. ■



JOHN ZHANG



CRISPR CAUTIONARY

A budding biologist put gene-drive ethics before experiments.

BY HEIDI LEDFORD

It was a trip to the Galapagos Islands at the age of ten that first whetted Kevin Esvelt’s appetite for tinkering with evolution. As he stood marvelling at the iguanas, birds and sheer diversity of the place that had inspired Charles Darwin, Esvelt vowed to understand evolution — and improve on it. “I wanted to learn more about how these creatures came to be,” he says. “And, frankly, I wanted to make more of my own.”

Today, Esvelt is still a precocious biologist. Less than a year after launching his lab at the Massachusetts Institute of Technology Media Lab in Cambridge, he has already made a name for himself as one of the pioneers of a controversial technique called a gene drive. His method harnesses CRISPR–Cas9 gene editing to circumvent evolution, forcing a gene to spread rapidly through a population. It could be used to wipe out mosquito-borne diseases such as malaria or eradicate invasive species. But it could also set off unintended ecological chain reactions, or be used to create a biological weapon.

The idea of CRISPR gene drives hit Esvelt when he was tinkering with the Cas9 enzyme in 2013. “I had one day of absolute, ecstatic glee: this is what’s going to let us get rid of malaria,” says Esvelt. “And then I thought, ‘Wait a minute.’”

Following that thought, Esvelt has worked to ensure that ethics comes before experiments. He first sounded the alarm in 2014, calling for public discussion about gene drives even before he had demonstrated that a CRISPR–Cas9 gene drive could work (K. A. Oye *et al.* *Science* **345**, 626–628 (2014); K. M. Esvelt *et al.* *eLife* **3**, e03401; 2014). Since then, he and his colleagues have shown how gene drives might be made safer, and how they could be reversed (J. E. DiCarlo *et al.* *Nature Biotechnol.* **33**, 1250–1255; 2015).

This year, his advocacy has begun to bear fruit. Researchers and policymakers worldwide have been discussing the technology, and a report from the US National Academies of Sciences, Engineering, and Medicine urged that gene-drive research proceed, but cautiously. Omar Akbari, who studies gene drives at the University of California, Riverside, believes Esvelt’s outreach has focused public attention — and attracted funding — for a nascent technology at just the right time. “I attribute that to Kevin,” says Akbari. “It’s difficult for a scientist to do what he’s done.” ■



PLANET HUNTER

An astronomer detected the nearest known planet outside the Solar System.

BY ALEXANDRA WITZE

Guillem Anglada-Escudé wasn't surprised early this year when evidence of an alien world rippled across his computer screen. He had been almost certain that an Earth-sized planet orbited Proxima Centauri, the star nearest the Sun at just 1.3 parsecs (4.2 light years) away.

To Anglada, an astronomer at Queen Mary University of London, the discovery came as more of a relief than a shock. He and his colleagues had been working feverishly to stake their claim in the competitive world of planet hunting, and the Proxima find confirmed that they were on the right path. "We made it," he says.

To the rest of the world, the discovery of the closest known exoplanet to Earth stoked the public imagination. It raised questions about whether life might exist in our cosmic backyard, and whether astronomers might be able to detect it.

These are the kinds of question that got Anglada into planet hunting in the first place. A science-fiction fan while growing up near Barcelona, Spain, he got his astronomical start doing data simulations for Gaia, a European Space Agency mission to map 1 billion stars. Later, he turned his data-crunching skills to exoplanets. He developed a method for extracting faint planetary signals from data gathered by the world's premier ground-based planet-hunting instrument, the High Accuracy Radial velocity Planet Searcher (HARPS) at

BRIAN DAVID STEVENS FOR NATURE

DIVERSITY TRAILBLAZER

A transgender physicist paved the way for greater acceptance of minority groups.

BY ELIZABETH GIBNEY

Physicists can be open to seeing the world in new ways, but they need to see the data first. This posed a problem for Elena Long, a nuclear physicist who has fought for her field to be more inclusive of people from sexual and gender minorities. "We didn't have any data, because people considered it too offensive to ask if we exist. It was a catch-22."

Long was one of the architects of a first-of-its-kind survey run by the American Physical Society (APS), charting the experiences of physicists who are lesbian, gay, bisexual, transgender or from another sexual or gender minority (LGBT).

The findings, presented to a packed room at the APS March meeting this year, were stark. Of the 324 scientists who responded, more than one in five reported having been excluded, intimidated or harassed at work in the previous year. Transgender physicists reported the highest incidence of discrimination. Long, who is transgender herself, was unsurprised. In 2009, she began work for her PhD at the Thomas Jefferson National Accelerator Facility in Newport News, Virginia, which lacked trans-inclusive employment protections and health-care benefits. She felt isolated without LGBT support networks. "I loved the work I was doing, and I loved the research. But it was rough," she says.

So she founded the LGBT+ Physicists support group and began pushing for greater recognition at the APS, which eventually created a committee to collect data on LGBT discrimination. Many physicists, she says, could not even understand the need for such a study. Thanks to Long and her colleagues, physics is emerging as exemplary in its approach to

Ones to watch

2017

CORI BARGMANN

SCIENCE PRESIDENT, CHAN ZUCKERBERG INITIATIVE

Bargmann is steering the research operations of a US\$3-billion effort by the philanthropic organization to cure, prevent or manage all disease by 2100.

ROBERT FEIDENHANS'L

CHAIRMAN, EUROPEAN XFEL

As the new head of the world's most powerful X-ray free-electron laser, Feidenhans'l will guide the €1.2-billion (US\$1.3-billion) facility during its ramp up to becoming fully operational by mid-year.

JEF BOEKE

CO-LEADER, HUMAN GENOME PROJECT—WRITE

Boeke is a director of an ambitious effort that is seeking to synthesize the human genome. He and others are already close to making a yeast genome.

WU WEIREN

CHIEF DESIGNER, CHINA LUNAR PROGRAMME

China's plans call for launching the Chang'e-5 mission in the latter half of 2017 to collect the first lunar rock samples to be brought back to Earth since the 1970s.

MARCIA MCNUTT

PRESIDENT, NATIONAL ACADEMY OF SCIENCES

With her experience in President Barack Obama's cabinet, McNutt will have a central role in representing US science during Donald Trump's presidency.

the European Southern Observatory in La Silla, Chile.

"Guillem has a natural talent of seeing the big picture where others see details," says Mikko Tuomi, an astronomer at the University of Hertfordshire in Hatfield, UK, and a collaborator of Anglada's.

But Anglada soon ran straight into high academic drama, tussling with other researchers over who deserved credit for discovering a planet bigger than Earth and smaller than Neptune orbiting the star Gliese 667C. "I could have left the field and done something else," he says. "But I took the decision of following it very aggressively."

He dived into HARPS data, publishing paper after paper on the planetary signals he discovered amid the background noise in the data. And then, as if to push back on all the secrecy and competition, Anglada launched a very public hunt for a planet orbiting Proxima.

He put together a team and got observing time on HARPS, as well as other telescopes that could double-check whether any promising evidence that they found was caused by stellar activity, which can mimic the signs of a planet (a problem that plagues many exoplanet claims). The researchers put nearly all their details on an outreach website and social-media accounts. Being so transparent "didn't seem dangerous at all," Anglada says. "We had a feeling nobody else would do this."

Within days, they confirmed that the planet was there; within weeks, they submitted a manuscript detailing their discovery. The planet, called

Proxima b, is at least 1.3 times the mass of Earth and orbits Proxima every 11.2 days.

Although it is close to its star, the world is within the 'habitable zone', where liquid water could exist on its surface. That makes it not only the closest known exoplanet of the 3,500-plus confirmed so far, but also a place where otherworldly life could thrive — a double bonus for researchers and science-fiction fans alike.

Just before the paper was published in *Nature* in August (G. Anglada-Escudé *et al.* *Nature* **536**, 437–440; 2016), Anglada e-mailed British sci-fi writer Stephen Baxter, author of the novel *Proxima* (Gollancz, 2013). They corresponded about what life might be like on a world with one hemisphere permanently facing a flaring star, as happens at Proxima.

People could eventually get a close-up look at Proxima b. The Breakthrough Starshot initiative aims to send fleets of tiny laser-propelled spacecraft to a nearby star, and it may target Proxima as its closest and best option.

Anglada's next step is to see whether Proxima b transits, or passes across the face of its star as seen from Earth. The chances are low, but if it does, then much more science can be gleaned when Proxima's light passes through the planet's atmosphere, if it has one.

And if the transit does not happen? Then Anglada may be off, to tease out some other signal of another world. ■

KANDICE CARTER

these issues, says Samuel Brinton, a board member of the society Out in Science, Technology, Engineering and Mathematics. "We are literally using their work to start changes for the better in multiple fields," he says. The APS accepted the recommendations made in the March report. And in August, a major APS division voted to move its 2018 meeting out of Charlotte, North Carolina, in response to a state law that forces people to use public toilets that match the gender they were assigned at birth.

Long has meanwhile won two young-scientist awards offered by her lab and become a co-leader on two new accelerator experiments. "I've known a lot of postdocs who've done voluntary work, and usually it compromises their science," says Karl Slifer, Long's postdoctoral supervisor at the University of New Hampshire in Durham. "I've never seen that in Elena." (Long attributes her strict time management to a computer program she designed that charts every hour of her day.)

Now Long is helping to set up an APS membership group focusing on diversity and inclusion, which she hopes will make it easier for scientists in other minority groups to flourish. "I'm sure there are other people facing problems in the field I never thought about," she says. "I don't want them to wait seven years to get to a place where they can have a voice." ■

ELENA LONG



COMMENT

POLICY A chief science adviser to the UK government reflects on 2016 **p.520**



CITIES The rich past and exciting future of urban farming **p.522**

PSYCHOLOGY Michael Lewis turns his novelistic brio to behavioural economics **p.523**

POLITICS Use epidemiology to model the spread of fake news and extreme views **p.525**

NICHOLAS KAMM/AFP/GETTY



Marine biologist Jane Lubchenco (right) has helped to launch training programmes that help scientists to engage with policymakers.

So you want to change the world?

In these tumultuous times, **Nancy Baron** urges scientists to speak from the heart to build public trust in research.

Fifteen years ago, when I began to work as a communications coach, few scientists felt that engaging outside academia was part of their job. Now, when I ask workshop participants to stand if they 'want to change the world', almost all rise to their feet. Often, they look around in astonishment to see so many peers standing with them.

I work largely with environmental scientists: ecologists, hydrologists, fisheries researchers and others documenting the escalating challenges that face our planet. Some are moved by the disappearance of the wildlife they study, or the brooding omnipresence of climate change. Others are dismayed by divisive political discourse and election results,

where evidence may have been devalued or dismissed. All are compelled to reach out.

This marks a radical shift. A decade ago I spent much of my time persuading scientists of the value of outreach to the public and policymakers. Today, most consider such communication crucial: they recognize that the publication of research in ►

► journals such as this one is not an end in itself, but a launch pad to further things. It is no longer a question of ‘Should I do this?’, but rather, ‘How do I do this?’.

Fortunately, there’s a quiet revolution happening, with workshops, guide books^{1,2} and much discussion on social media about effective communication. This revolution is buttressed by social-science research on the benefits of turning stereotypically stiff and formal scientists into warm, accessible people. And it is supported by a growing network of scientists who mentor each other’s outreach. There has been a shift from peer pressure to keep quiet, to peer support to speak up.

This year more than ever it has become clear that the revolution can be kept quiet no longer. It is time for scientists to take social responsibility and to be recognized and rewarded for doing so.

JUDGEMENT DAY

Fear of being judged by one’s peers still haunts many academics. No matter who they are talking to, scientists often fret about how they might sound to their colleagues instead of concentrating on the audience at hand. They also worry about being subjected to the sort of abuse hurled at climate scientist Michael Mann. His famed hockey-stick graph, published in 1999, reconstructed a temperature record over the past 1,000 years. His resolve to speak out about his findings plunged him into a maelstrom of hacked e-mails and smears by climate-change deniers.

But a 2014 study³ showed that scientists might risk less criticism than they think — at least from each other. A survey carried out at nine ecological and environmental conferences found that, to a surprising degree, scientists wanted to engage with the public and policymakers more. Many respondents believed that scientists should interpret and even advocate for the use of their science. A decade earlier, a similar study showed that ecological scientists strongly disagreed with the idea of active advocacy⁴.

The question of how scientists should engage is a deeply personal one⁵. Nevertheless, there is a gap between desire and action. In the 2014 survey, the main barriers were a self-perceived lack of competence at navigating the science-policy interface, as well as past negative experiences and institutional norms that did not support them. Lack of time and resources, which often spring to mind first, were lesser factors. The survey found that the more scientists were aware of how their work fits into the policy landscape, the more likely they were to get involved.

‘HOW TO’

There are organizations that can help. Two examples are: COMPASS, a non-profit, non-advocacy programme (for which I work) that helps scientists to engage effectively



Strident campaigning is just one way to stand up for science. Warm and well-timed conversations with key players are crucial too.

in the public discourse about the environment; and the Leopold Leadership Program, now a part of the Stanford Woods Institute for the Environment in California. Both were launched under the leadership of Jane Lubchenco, a marine ecologist at Oregon State University in Corvallis (before she became chief of the National Oceanic and Atmospheric Administration), and others. These pioneering programmes were created to, as Lubchenco has said, aid “faster and more effective transmission of new and existing knowledge to policy and decision-makers and better communication of this knowledge to the public”.

Since 2000, my colleagues and I at COMPASS have worked with thousands of environmental scientists. Over the years we have learned that teaching them communication skills is necessary — but not sufficient. Between the desire to communicate and actual engagement, there is a ‘valley of death’ that can stop scientists from trying out their skills in the real world. They need guidance on how to bridge that gulf.

The first step is to examine the big picture and determine what your role in a conversation might be. What do you uniquely bring to this issue? Next, analyse potential opportunities: identify the players, including the decision-makers and stakeholders, and determine what they need — and when. This is the key to making advice relevant and timely. Professional societies, government-relations departments at some universities, and organizations such as

COMPASS can help to broker connections. So, too, can other scientists.


LINKED IN

Once a contact is made, a single successful interaction often creates new opportunities. That first action might be an opinion piece in a local paper, speaking to community members, or an e-mail to someone in local government that meets with a positive response. Even just preparing properly for media attention to an academic publication can launch conversations with decision-makers that take on a life of their own.

This is what happened to Jenna Jambeck, who studies waste-management engineering at the University of Georgia in Athens. In 2015, a paper by Jambeck’s team about international data on how much plastic debris enters the ocean each year was published in *Science*⁶. Because her working group of scientists was at the National Center for Ecological Analysis and Synthesis, where COMPASS has a staff presence, she asked for guidance. We helped whittle her 26 statistics down to a few compelling points: “Humans dump 8 million tonnes of plastics into the oceans each year. That’s five grocery bags of plastic for every foot of coastline in the world,” for instance. When the media flood came, she was prepared.

This launched Jambeck’s role in the political conversation about plastic waste. Since then, she has become a sought-after expert at congressional hearings and around the globe. Her science is important, but it is

DOMINIC FAURE/KEystone/AP



also her conversation that makes her so successful. Her voice resonates with people because of her optimism that this problem can be solved, along with her willingness to show her passion for recycling and waste management. One of her best lines is that she met her husband at the landfill.

Jambeck's friendly tone matters. A study this year revealed that scientists most prioritize communication designed to educate and defend against misinformation, and that they least prioritize communication designed to build trust and resonance with the public⁷. Yet research shows that people's willingness to listen is linked to how likeable, warm and authentic they find the speaker⁸. Building trust requires a human touch.

Ken Caldeira, an atmospheric scientist at the Carnegie Institution for Science in Stanford, gave a lecture on World Oceans Day to leaders of the United Nations Educational, Scientific and Cultural Organization in Paris this year. He presented the perils of acidifying oceans but reframed them as a question: how can we help the oceans help us? This empathetic approach resonated. Focusing on solutions, and using inclusive language such as 'we' and 'us' when talking to audiences makes scientists — and their findings — more approachable. Caldeira has also embraced social media. Twitter, Facebook and other social platforms are key to shifting public opinion, as events this year have shown. Scientists should join these conversations to extend their reach. Sometimes, despite doing the right work and reaching out to the right people, the science does not prevail. Jonathan Moore, a coastal scientist at Simon Fraser University in Vancouver, Canada, spent four years tallying up the social and environmental havoc that would be caused by a liquid natural-gas terminal on the northern coast of British Columbia: it would disrupt habitats that support salmon relied on by 11 First Nations groups over an area the size of Switzerland, his team found. Although Moore and his colleagues have published in the peer-reviewed literature⁹, been covered by hundreds of media outlets and met with communities and policymakers, the terminal was granted government approval by a regulatory review process that many Canadian scientists find lacking.

SUPPORT NETWORK

Public engagement takes perseverance and courage. It also needs emotional support. It's a long haul, and hard to do alone. In 2014, Moore was a member of the first cohort of the Wilburforce Fellowship, run by a charitable foundation in Seattle, Washington. The year-long programme is designed to provide leadership and communications training to scientists at all career stages to form lasting networks of support across Canada, the United States and Mexico. Initial training

lasted only a week, but fellows continue to share successes, failures and advice for resilience through social media and e-mail. At conferences, they stay up late discussing how to advance their goals and offer each other encouragement.

Some of the earliest fellowships, including the Leopold programme, were intended specifically for tenured academics, because outreach was seen as a career risk. This is changing. The next generation of young scientists is most eager to change the world, and is stepping up to do so.

Canada's government, elected last year in place of an administration that was accused of muzzling government scientists, has promised to usher in an era of evidence-based decision-making. More than 1,800 early-career researchers have contributed recommendations on how to rebuild confidence in environmental assessments and regulatory reviews — so far with mixed results (see youngresearchersopenletter.org).

Sally Otto, an evolutionary biologist at the University of British Columbia in Vancouver, directs the Liber Ero (Latin for 'I will be free') fellowship. It funds early-career scientists to do applied research on important conservation issues and provides training in public and political outreach. Otto also donated some of her 2011 MacArthur 'genius grant' to launch a policy fellowship that embeds scientists in government in Ottawa. The scheme is run by Mitacs, a Canadian non-profit organization that specializes in partnerships and placements between academic and non-academic institutions.

Young scientists are under massive pressure to win grants and publish. So a growing cadre of senior scientists is instigating fellowship programmes that provide early-career researchers with communication skills and connections, and in some cases also funding.

Such efforts are changing the academic reward systems, albeit too slowly. Several institutions now have tenure packages that recognize communication and outreach, along with conventional measures of publication and teaching success. Lisa Graulich, dean of the College of the Environment at the University of Washington in Seattle, has expanded her definition of impact to include public engagement, which she considers a logical product of engaged scholarship. Some argue that engagement, although harder to measure than citations, is a better proxy for academic success¹⁰.

Training on how to extend into the public realm and understand the workings of government shouldn't be available only through boutique fellowships or under the wings of

a few motivated senior faculty members. It should be a part of every young scientist's education. In 2014, COMPASS led a working group funded by the US National Science Foundation (Building Systemic Communication Capacity for Next Generation Scientists) to assess the science-communication workshops and training available to US graduate students in science, technology, engineering and mathematics. We provided recommendations for integrating science-communication skills into graduate education (see compassblogs.org/gradscicomm). So far, these early efforts, in need of funding, have languished.

SPEAK OUT

In these uncertain times, the voices of scientists are more important than ever. Efforts should not only target political leaders, but also aim to create a groundswell of public support. Ultimately, leaders must listen to their constituents. This is a time for scientists to double down and launch a ground campaign for the hearts and minds of the public.

Society needs to hear from those who can explain empirical evidence in a way that resonates with people's values, whatever they may be. We all need to be more open-minded and inclusive — and we need to muster the courage to speak from the heart and learn to listen with empathy.

My experience is that scientists can emerge as powerful agents of change. By building capacity, collaboration and confidence among researchers, we can bolster public engagement, inform decision-making and inspire society to forge a better future. ■

Nancy Baron is science outreach director for COMPASS, a public-engagement organization for scientists, and author of *Escape from the Ivory Tower: A Guide to Making your Science Matter*. She is based at the National Centre of Ecological Analysis and Synthesis in Santa Barbara, California, USA.
e-mail: nbaron@compassonline.org

1. Baron, N. *Escape from the Ivory Tower: A Guide to Making Your Science Matter* (Island, 2010).
2. Badgett, M. V. L. *The Public Professor: How to Use Your Research to Change the World* (NYU Press, 2016).
3. Singh, G. G. et al. *Front Ecol. Environ.* **12**, 161–166 (2014).
4. Lach, D., List, P., Steel, B. & Shindler, B. *BioScience* **53**, 170–178 (2003).
5. Pielke, R. A. *The Honest Broker: Making Sense of Science in Policy and Politics* (Cambridge Univ. Press, 2007).
6. Jambeck, J. et al. *Science* **347**, 768–771 (2015).
7. Dudo, A. & Besley, J. C. *PLoS ONE* **11**, e0148867 (2016).
8. Fiske, S. T., Dupree, C. *Proc. Natl Acad. Sci. USA* **111**, 13593–13597 (2014).
9. Moore, J. W. et al. *Mar. Ecol. Prog. Series* **559**, 201–215 (2016).
10. Hoffman, A. *Academic Engagement in Public and Political Discourse* (Univ. Michigan Library, 2015).

Supplementary information accompanies this article online at go.nature.com/2hvnspi.



ILLUSTRATION BY DAVID PARKINS

Take the long view

At the end of a difficult year for evidence, **Ian L. Boyd**, a chief scientific adviser to the UK government, draws lessons for making research more relevant.

The post-truth politics of 2016 has made it a difficult year for those of us who like to see decisions guided by evidence. As a chief scientific adviser to the UK government, I would caution against any disengagement by the science community from policymaking. Science remains an integral part of the processes of government, and its outputs are increasingly needed.

In the areas for which I have responsibilities — food and environment — there is an obvious need to maintain momentum in finding solutions, such as for mitigating climate change. The deeper we dig, the more we understand the important role of the environment for human health and welfare, including for inequality. Achieving the United Nations' Sustainable Development Goals requires a strong and unified approach by the science community, irrespective of populist politicians and policies.

In reality, government is mostly preoccupied with reacting to acute events. Strategy swings with the political whim, the urgent crowds out the important and policymaking can become a displacement activity — a survival strategy that creates the impression

of progress against a background of intractability. As a result, it is often only scientists in and around government who carry the baton for confronting the brutal realities of environmental challenges.

There is a danger that, in the maelstrom of day-to-day policy delivery, scientific input will be reduced to a mere technical instrument. To be more involved in creating policy, scientists need to focus on different research priorities from those normally seen as important to their careers.

HARDEST QUESTIONS

This different posture is both structural and intellectual. Structurally, to become trusted components of the policy process, scientists have to develop a heightened appreciation of how government works. Working in government-related science needs to be valued as equal in importance to working in academia. (Having done both, I can say that this is not currently the case.) The constraints on how government scientists behave, and on what they can and cannot say at any particular time, need to be appreciated much more sharply by scientific colleagues and press intermediaries.

Intellectually, we need to frame the challenges faced by government using language that reveals their intrinsic value to scholarship and academic progress. We have perceptions of what 'excellent' science looks like, but often, almost by definition, this does not include many of the local, urgent and multidisciplinary questions being addressed in government.

Yet governments tackle some of the most difficult questions facing people and the planet — from what particular price for carbon might affect employment to what portion of the health budget should be spent on prevention. Costing environmental degradation in to the decisions made in everyday life stands out as one of the greatest challenges. Feigning to ignore these, or focusing on only one component of them to the detriment of building a broader understanding of how they might be solved, happens too often in academic science.

SYSTEM CHANGE

In practice, governments manage systems — farming, say, or transport. These include many interacting processes and

actors, natural and social. In general, governments have little control over the components of these systems. What they most need scientists' help with is understanding which parts can be managed, and studying the behaviour of whole systems in response to those changes, often through data about system indicators and research into their dynamics.

For example, scientific advice about fisheries tends to wrongly assume that fisheries management is about managing fish, when we can normally only manage those who fish. Similarly, the solution to the bovine tuberculosis epidemic in Britain is more about managing farmer behaviour than it is about applying well-trying epidemiological solutions.

The contribution made by supply-side innovation — that is, inventing new materials, devices or structures, or probing the complexity of nature — is undoubtedly a good thing. It is often touted by politicians as the main way in which research adds value in civil society. As scientists, we often acquiesce to this linear view because it is the route through which money tends to flow. Politicians like it because it supports market activity and economic growth, and gives people more of what they want.

However, for most policies to work, equal strides are needed in demand-side innovation (research that answers specific, pragmatic questions). In food and environment, this often involves increasing system efficiency and reducing demand.

For example, many countries have now legislated for the use of low-energy lighting, stimulated by supply-side innovation that increased light-bulb efficiency by at least an order of magnitude. But a lack of simultaneous demand-side innovation, such as through behaviour change, has meant that the overall power usage has continued to rise — we just use more lights. The same goes for the fuel efficiency of cars. Arguably, by focusing on one part of the problem, science has only added to its intractability.

PROBLEM-SOLVING

Tackling many of the environmental grand challenges will need large-scale investments in system models. Climate science has emerged from the need to forecast weather and so provides a template for how this could be done. It involves high levels of organizational design, including a global environmental-observation network providing large data flows linked to ocean and atmospheric models run on high-performance computers.

Many other areas of environmental science — air and water, food, waste, and biodiversity — need a similar scale of effort and investment. The data flows

from observational networks are emerging, but there is insufficient coordination of system-model development to capitalize on these data.

Achieving this requires changes in policy for science and social science to incentivize the research community to participate in government. This means reassessing the relative value of game-changing discoveries (supply-side innovation) over the organizational, system-based solutions that are needed (demand-side innovation).

If researchers want to play their part in solving major problems, such as decoupling economic growth from resource consumption, they need to change their focus. This requires greater prioritization of behavioural and operational research, a discipline that gets scant coverage in academic circles but which encompasses systems analysis and modelling. It also requires greater value to be placed on synthesis as a tool in discovery because of its power to describe system-level behaviour.

Often it is the simple solutions applied well that make the difference rather than new technologies.

This is not the world of the laboratory bench or the individual theoretician. It is one in which system models are being continually refined on the basis of big, open data about the system's state and its responses. This will blur the boundaries between experimentalists and those who run the policies — because a policy becomes a hypothesis. And it will turn science back from the path of being perceived as an irrelevant domain of the intellectual elite. Recent growth in anti-science views on both sides of the Atlantic suggests that this change is imperative.

WISH LIST

So, what are the really big systems challenges in my areas of responsibility? First, in my view, we need to know much more about the future of resources. Raw materials drive the global economy; if they cannot be grown, they need to be mined. In response, we have invented solutions such as the circular economy. Although no one doubts the wisdom of driving up the productivity associated with the materials already in the economy, their reuse may divert attention from difficult decisions about reduction. Better systems-based models of resources and materials are needed to help frame the policy options.

Similarly, we need to know the level of assurance of our worldwide food supply. It is difficult for policymakers to estimate how much reserve is needed to create

resilience to different kinds of shocks, natural or human-made. Most governments currently leave this crucial function to the market, but is this wise? Little is known, beyond what equilibrium economic models tell us, of the stability and resilience of food-supply networks. Many human-made and natural networks show nonlinear behaviour and have a capacity to reach a tipping point. Could this happen to global food supplies?

An extension of this question concerns the future of livestock. This has very low levels of material efficiency, so shifting away from livestock production might simultaneously address concerns about future food supplies and resources. Although livestock production can be the best use of marginal land and is important in some developing countries, it is also a significant contributor to greenhouse gases, exacerbates the problems of antimicrobial resistance, causes pollution, increases the risks from diseases that are spread from animals to humans, and drives the destruction of tropical forest. Furthermore, current levels of meat consumption in the developed world are unhealthy. Are there different systemic solutions to meat production and consumption that address these kinds of problem?

There are other priorities, of course. But in the interests of focusing on finding simple solutions and applying these well to achieve maximum effect, these examples could address many of the large-scale and long-term environmental challenges facing the planet. Intellectual resources need to be deployed where they will make the biggest difference, and this requires leadership and vision.

Politicians who are willing to listen may say this is all too difficult. However, scientific leadership can help policies and the systems they are designed to influence to evolve together. Pointing to small changes in key variables and introducing changes incrementally can have big effects over time.

This is illustrated by how some countries, including Britain and several other European nations, are on track to eliminate sending waste to landfill. Incremented taxation of landfill waste has changed behaviours and encouraged investment in recycling and reuse. A simple tax applied in the right place and appropriately scaled has shifted the whole system state. It has changed behaviours without stranding assets and, importantly from a political perspective, it has not upset the electorate. ■

Ian L. Boyd is chief scientific adviser in the UK Department for Environment, Food and Rural Affairs and professor of biology at the University of St Andrews, UK.
e-mail: ilb@st-andrews.ac.uk



CYRUS DOWLATSHAHI

Brooklyn Grange, a 1-hectare organic farm that spans two rooftops in New York City.

AGRICULTURE

Sowing the city

As holiday feasts begin, **Laura Lawson** surveys the fruitful history of urban farming.

The floating gardens of Aztec Mexico; the Versailles kitchen garden that supplied the table of Louis XIV of France; hydroponic vertical farming in Japan. Throughout history, agriculture has been integrated into cities in various forms, moulded by environmental conditions, design intent and technological and agromonic innovation. Today, an explosion in studies on urban agriculture is broadening perspectives. It's illuminating the view from Africa and Latin America in Luc Mougeot's 2005 *Agropolis* (Earthscan), the varied historical contexts in Dorothée Imbert's edited *Food and the City* (Harvard Univ. Press, 2015) and the design propositions in André Viljoen's 2005 *CPULs: Continuous Productive Urban Landscapes* (Routledge). These books not only normalize urban farming, but also address how it can be adapted to tackle food security.

According to the 1996 report *Urban Agriculture* by the United Nations Development Programme, an estimated 800 million urban dwellers worldwide grow food and raise livestock. However, many people still

ask whether urban agriculture can grow enough food to be useful. The assumption is that farming is best as a rural practice, with extensive land and trained farmers to maximize productivity. Many others are increasingly concerned about the impacts of the agro-food system on health and the environment, and look to the different possibilities of urban agriculture — addressing nutrition, economic development, community activism and environmental awareness.

Urban farming is, in fact, rarely bracketed with the agro-food system, but it is crucial to initiatives that address food access and affordability. It also gives farmers the flexibility to choose crops on the basis of genetic diversity or cultural preferences — such as 'heritage' vegetables or varieties valued by recent immigrants. It is not a variation on rural agriculture, but a distinct practice.

GROWING TOGETHER

Farming has been part of the structure of cities for millennia, with farmers supplying urban markets and repurposing waste such as manure (see 'Capitals of cultivation'). In

The Economy of Cities (Random House, 1969), urban theorist Jane Jacobs (A. Williams *Nature* 537, 614–615; 2016) argued that agriculture did not precede early cities: instead, cities inspired agriculture through their centrality to trade and capacity for innovation. Archaeological evidence from France to Iran — of canal and irrigation systems, terracing and walled enclosures — reveals the large investment often needed to sustain urban productivity.

Colonization plans such as Spain's Law of the Indies — which held for 400 years in the Americas and the Philippines — placed agriculture close to cities such as Santo Domingo to ease access to crops and animals, provide waste management and protect food supply. In 1683, William Penn, founder of Pennsylvania, planned Philadelphia as a "green country town" where productive orchards and gardens would ensure fresh air and reduce the risk of fire. Later, Ebenezer Howard's British 'garden cities' incorporated allotments and farms in 'green belts', starting with Letchworth at the turn of the twentieth century. In 1928, the German landscape architect Leberecht Migge

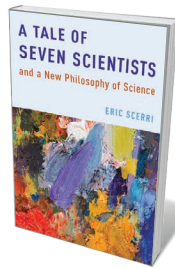


proposed to integrate allotment gardens into public housing developments around an expanding Frankfurt. And US architect Frank Lloyd Wright planned a decentralized settlement (never realized) called Broadacre City, in which each family was allotted almost half a hectare for food production.

Whether informal or planned, agriculture has had to adapt to urban expansion. In some cases, this has led to intensive, innovative cultivation to increase yields and support higher-value goods. From the twelfth to the nineteenth centuries, farmers in the Marais district of Paris transformed swampland into productive gardens and manipulated microclimates to grow peas, artichokes, figs and apricots. They planted densely in raised beds nourished with (and kept warm by) horse manure, and used cold frames and cloche coverings to extend the growing season, as well as trellis supports to encourage fruit yields. Yet even with innovation and intensification, rising land costs and improved transport often led farmers in Paris and elsewhere to relocate to otherwise unbuildable sites, or out into the urban periphery.

City farming has also been a way to relieve poverty. The nineteenth and twentieth centuries saw English urban allotments, French *jardins familiaux* and *jardins ouvriers*, German *kleingärten* and, in US cities such as New York, vacant-lot cultivation associations and relief gardens that provided land and training as both direct aid and social reform. Gardening was about food, but also about tradition, culture, education and morale, for instance harnessing the know-how of immigrants. These programmes laid the way for domestic food production in Australia, Europe and North America during the two world wars, when hundreds of thousands of civilians grew their own food so that commercial crops could serve the war effort. Second ▶

Books in brief



A Tale of Seven Scientists and a New Philosophy of Science

Eric Scerri OXFORD UNIVERSITY PRESS (2016)

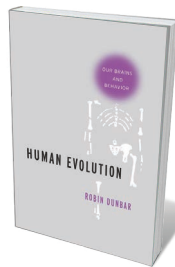
Is scientific discovery all about starry intellects and eureka moments? Philosopher of science Eric Scerri asserts otherwise in this thoughtful treatise on research as evolution, not revolution — collective, piecemeal endeavour rather than heroic act. Scerri spotlights seven ‘unknowns’ who, pivoting around the work of Niels Bohr and Dmitri Mendeleev, helped to unravel atomic structure. So physicist John Nicholson’s concept of the quantization of angular momentum informed Bohr’s quantum theory of the atom, and the amateur Anton van den Broek pioneered the idea of atomic number.



The European Research Council

Thomas König POLITY (2016)

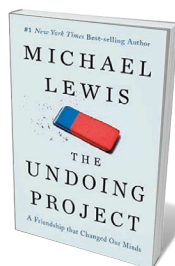
A beacon of science policy and research funding, the European Research Council (ERC) is about to celebrate its first decade. In this first comprehensive history, Thomas König — former scientific adviser to ERC president Helga Nowotny — offers a multifaceted perspective. Behind the ERC’s success in maintaining excellence in ‘frontier’ research, he reveals, a welter of complex negotiations have played out in the politicized current of its framework funding programme, Horizon 2020. A story of big scientific personalities and struggles for autonomy and accountability in the charged space between policy and science.



Human Evolution: Our Brains and Behavior

Robin Dunbar OXFORD UNIVERSITY PRESS (2016)

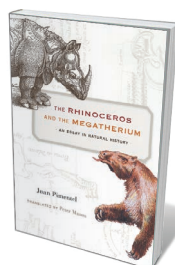
“Stones and bones” dominate the human-evolution story, notes evolutionary anthropologist Robin Dunbar. In this exemplary study, he plunges instead into the “murky, unseen social world” during the 6 million to 8 million years since the hominin lineage diverged from that of other African great apes. Drawing on research from the British Academy’s From Lucy to Language project (H. Gintis *Nature* **509**, 284–285; 2014), Dunbar traces shifts in biology, genetics and neurology over that key period. It’s a compelling journey into human nature, from the roots of our sociality to the rise of storytelling.



The Undoing Project: A Friendship That Changed Our Minds

Michael Lewis W. W. NORTON (2016)

Soon after Michael Lewis published *Moneyball* (W. W. Norton, 2003), his best-seller on the metrics of sports recruitment, he found that the concepts were not original. Psychologists Daniel Kahneman and Amos Tversky got there decades before, codifying the systematic bias in our decision-making when faced with uncertainty. Lewis tells the story of this rare scientific collaboration and its impact on behavioural economics with novelistic brio, tracing the dual evolution of Kahneman and Tversky’s intense relationship and their research all the way to a denouement of human sorrow and Nobel glory.



The Rhinoceros and the Megatherium: An Essay in Natural History

Juan Pimentel; transl. Peter Mason HARVARD UNIVERSITY PRESS (2017)

In this subtly discursive study, historian of science Juan Pimentel looks at two animals that profoundly marked science, yet were “imagined without being seen”. One was the rhinoceros famously hypothesized by Albrecht Dürer in a 1515 woodcut. The other was a *Megatherium* fossil discovered in 1788 that, rendered in an engraving, allowed comparative anatomist Georges Cuvier to identify it as a giant sloth. Pimentel’s inspired pairing limns how image and imagination shape our understanding of nature. *Barbara Kiser*



Aztecs farming on an artificial islands called a chinampa in Tenochtitlan.

CAPITALS OF CULTIVATION

Agriculture has been part of the urban infrastructure since cities first began.

Twelfth century: Farmers in the Marais, Paris, pioneer intensive cultivation, using horse manure for fertilizer and warmth.

Fourteenth to sixteenth centuries: The chinampas of Tenochtitlan, Mexico — artificial islands built out into lakes — generate up to two-thirds of the Aztec capital's food.

1683: William Penn plans the city of Philadelphia in Pennsylvania colony as a utopian "green country town" with productive orchards and fields.



1940s: Garden 'armies' in Australia, Britain and the United States boost domestic self-sufficiency, leaving commercial crops to feed the war front.

1970s: Amid recession, community gardening takes root in US cities from New York to San Francisco, transforming vacant lots into cropland.

1980s: With the loss of Soviet trade subsidies, Havana reinvents urban farming as intensive 'organoponics', and reintroduces oxen for ploughing.



1990s: Aeroponics and LED lighting allow hyperflexibility — vertical fields, indoor farming and commercial urban agriculture.

2000s: To support China's urban expansion, green belts such as that around Shanghai grow around half of the city's vegetables.

2016: The US Department of Agriculture unveils the *Urban Agriculture Tool Kit*, offering technical advice for city farming.

► World War 'victory gardens' in Britain, for instance, saw everything from swathes of London's Hyde Park to railway embankments do duty as cropland. Community-garden activism in the 1970s and 1980s built on the legacy of such programmes but focused on community empowerment. Movers and shakers emerged, such as Liz Christy, first director of the Council on the Environment of New York City's Open Space Greening Program. And the movement led to neighbourhood gardens and city-wide advocacy organizations including Boston Urban Gardeners in Massachusetts.

A MOVEABLE FEAST

Over time, unique tensions have emerged around urban agriculture. Its mess, noise and perceived unruliness can vie with city authorities' urge towards order and cleanliness. Animal husbandry has been a flashpoint. Millions of urban dwellers in the developing world raise everything from chickens to camels, but this form of farming was an early casualty in the West. Philadelphia's free-ranging pigs, for instance, succumbed to early-twentieth-century legislation and the closure of urban livestock markets.

The dirty reputation of farming persists in some Western cities. With a globalized food system capable of propelling asparagus from Peru into British supermarkets, food planning and policy in Western cities tend to focus on trucking and warehousing, leaving officials prone to viewing localized production as a regressive, temporary use of urban space. Some urban-farm designers attempt to neaten up the image with colourful geometric plant displays next to street cafes and on rooftops. But they risk promoting style over substance: such plans rarely acknowledge the realities of cultivation, such as seasonality and the need for fallow beds.

Thus urban agriculture has become a moveable feast — opportunistic or mandated, marginal or essential. Since the 1990s, initiatives by bodies such as the Food and Agriculture Organization of the United Nations have encouraged policies, pilot projects and micro-lending to support farming for food security in many poorer African and Asian cities. And in socio-economically deprived postindustrial regions such as Germany's Ruhr valley and the Rust Belt in the US northeast and midwest, urban agriculture is seen as generating economic activity and enabling access to better food. Growing Power, an organization founded by former basketball star Will Allen in Milwaukee, Wisconsin, provides training in closed-loop 'aquaponics' among other urban production practices across the United States (see go.nature.com/2gzjh4p).

Elsewhere, Chinese megacities such as Shanghai are supporting agricultural green belts to retain local access to fresh vegetables. Cuba, long separated from global markets,

produces roughly 60% of its vegetables in towns and cities, with remarkable success based on traditional practices such as reintroducing oxen for ploughing. In 2011, San Francisco, California, passed an urban-agriculture ordinance that made farming an allowable use of land and source of saleable food. Agriculture is also included in the sustainability plans of cities such as Seattle, Washington, to reduce its carbon footprint and green the infrastructure. And housing developments including Serenbe outside Atlanta, Georgia, are embracing community-supported agriculture, farmers' markets and farm-to-table cuisine as a 'lifestyle choice'.

In tandem with the push for sustainable business and circular economies (B. Kiser *et al. Nature* 531, 443–446; 2016), high-tech options are emerging. Vertical urban farms, hydroponic greenhouses and aeroponic factories offer intensive production to offset high real-estate costs. AeroFarms in Newark, New Jersey, perfected its aeroponic salad-

"Urban agriculture will be there when new challenges arise."

growing process at a local school in 2011 and now anticipates harvests of more than 4 million kilograms per year. But these capital-intensive projects demand business savvy, and critics argue that their presence can lead to older models such as community gardens being undervalued.

Agricultural urbanism is entering a new phase as a framework to address community cohesion and food access. From Shanghai to Detroit, advocates are mapping the urban-agriculture landscape — highlighting the existence of vacant lots and 'food desert' neighbourhoods ripe for transformation. Often, this enables farmers to network, discuss shared concerns and advocate. A model is New York City's Five Borough Farm, a project of the Design Trust for Public Space. Here, site documentation, metrics development and proposals for supportive policies and practices are managed collectively.

And the future? Urban agriculture offers promise for coping with climate change. Eschewing reliance on vulnerable transport connections, experimenting with seasonality and crop selection, and strengthening community ties will help both mitigation and adaptation. Urban agriculture will be there when new challenges arise, and will continue to evolve as it responds to key issues that shape our cities. ■

Laura Lawson is dean of agriculture and urban programmes at Rutgers University in New Brunswick, New Jersey. Her books include *City Bountiful* and *Greening Cities, Growing Communities* (co-authored with Jeffrey Hou and Julie Johnson). e-mail: ljlawson@sebs.rutgers.edu

Correspondence

Study epidemiology of fake news

The results of the 2016 US presidential election and the UK vote to leave the European Union (Brexit) have raised questions about the influence of fake online news and social-media 'echo chambers' (see also P. Williamson *Nature* **540**, 171; 2016). The propagation of such information through social networks bears many similarities to the evolution and transmission of infectious diseases. Analysis of transmission dynamics could therefore provide insight into how misinformation spreads and competes online.

For example, disease strains can evolve and compete in a host population, much like rumours, and infections and opinions are both shaped by social contacts. Modelling of competing disease strains indicates that, as contacts become more localized, the diversity of circulating strains can increase (see C. O'F. Buckee *et al. Proc. Natl Acad. Sci. USA* **101**, 10839–10844; 2004). Network structure can also suppress the invasion of new disease strains (see G. E. Leventhal *et al. Nature Commun.* **6**, 6101; 2015).

As more people turn to social networks as a primary news source, transmission models combined with appropriate data could help in exploring the dynamics of this new media landscape.

Adam Kucharski *London School of Hygiene & Tropical Medicine, UK.*
adam.kucharski@lshtm.ac.uk

Be wary of 'ethical' artificial intelligence

Jim Davies's suggestion that we programme ethics into artificial intelligence meta-systems as a safeguard could well backfire — by compromising our abilities to judge ethical implications (*Nature* **538**, 291; 2016).

In an earlier version of the future, robot lawnmowers and

kitchen appliances promised us more leisure time. We now face the spectre of mass human displacement from a consumption-based economy by equipment that can do things much more efficiently than people can.

The 'age of information' promised global connectivity, but this has wrought distraction to a point at which only lurid excesses can focus our undivided attention on the society to which we all belong.

And as computer-generated imagery colonizes our imaginations, many are barely swayed by real violence (the wanton destruction of Syrian cities comes to mind). There is even evidence that video gaming driven by computer-generated imagery can alter a player's perception of acceleration and gravity (see, for example, A. B. Ortiz de Gortari and M. D. Griffiths *Int. J. Hum. Comput. Interact.* **30**, 95–105; 2014) — compromising their decision-making skills in a world where real physics is the law. Such trends don't bode well for 'ethical' computers.

Michael Stocker *Ocean Conservation Research, Lagunitas, California, USA.*
mstocker@ocr.org

New contender for most lethal animal

Of some 3,500 species of mosquito, those of the genus *Anopheles* are widely considered to be the most dangerous because they transmit malaria. Malaria is decreasing, however, and other mosquito-borne diseases, such as dengue, chikungunya, Zika and yellow fever, are increasing (S. V. Mayer *et al. Acta Tropica* **166**, 155–163; 2017). The mosquito *Aedes aegypti*, the primary carrier of these viruses, now constitutes an even greater threat (see also *Nature* **539**, 17–18; 2016, and S. F. Dowell *et al. Nature* **540**, 189–191; 2016). *Aedes* mosquitoes have been

transferring these viruses among African primates for millennia. One African primate (*Homo sapiens*) and one African *Aedes* (*A. aegypti*) have spread from Africa. Viruses adapted to both have spread with them. Yellow fever hit the developed world in the seventeenth century, dengue in the nineteenth, chikungunya in the twentieth, and now Zika in the twenty-first.

Scientists studying mosquito-borne viruses have catalogued hundreds more lurking in Africa. The world needs to take notice before these take hold and spread further. As Pliny the Elder (AD 23–79) wrote, "*Ex Africa semper aliquid novi*" ('there is always something new coming out of Africa').

Jeffrey R. Powell *Yale University, New Haven, Connecticut, USA.*
jeffrey.powell@yale.edu

Journals, agree on manuscript format

An 'incorrectly' formatted manuscript submission risks immediate bounceback by the authors' chosen journal, irrespective of the value of its content. In my view, it would save time and frustration if the scientific community could agree on a uniform style for all journals.

There is no inherent advantage in customized formatting of references, for example, whether cited as F. R. Smith, P. Y. Young and G. T. Jones *J. Interest. Sci.* 2016, **85**, 6700–6782, or as Smith, FR, Young, PY, Jones, GT (2016) *J. Interest. Sci.* **85**: 6700–6782, or using other arbitrary variants in style and positioning of initials, year of publication and page span.

Research papers in the natural sciences are typically presented under the headings Abstract, Introduction, Methodology, Results, Discussion, Conclusions. Some journals do not use Abstract or Introduction headings; some put the Methodology section after the rest. No journal so far puts the title at the end of the paper.

Journals presumably insist on individual formatting styles as a distinguishing feature. I see no scientific merit in doing so. Cosmetic treatments should instead be reserved for enhancing the clarity of a manuscript's content.

Quanmin Guo *University of Birmingham, UK.*
q.guo@bham.ac.uk

Vulture restaurants cheat ecosystems

Vulture 'restaurants' across southern Europe are serving up carcasses in an attempt to rescue these endangered birds. We contend that such outlets are no true replacement for the naturally random food pulses associated with wildlife carrion.

When food is only randomly available, it supports the foraging of hundreds of invertebrate and vertebrate scavenger species, promoting co-existence of multiple species and thus ecosystem balance. By contrast, a predictable food supply tends to benefit only selected species (see, for example, A. Cortés-Avizanda *et al. Front. Ecol. Environ.* **14**, 191–199; 2016) — griffon vultures (*Gyps fulvus*) in this case.

Furthermore, large predators are now expanding in Europe, contributing to the rewilding of many landscapes. Predation helps to stabilize food chains by buffering oscillations in the availability of carrion, which are linked to climate events and disease epidemics. Scavenger species that recover as a result of such ecological rewilding boost ecosystem functions and services derived from their feeding behaviour, for example by reducing the spread of disease.

We should stop providing services to nature through vulture restaurants and allow nature to provide services to us.
Ainara Cortés-Avizanda, Henrique Miguel Pereira *CIBIO-InBIO, University of Porto, Vairão, Portugal.*
hpereira@idiv.de

METASTASIS

Pathways of parallel progression

Two studies in mice identify mechanisms by which tumour cells disseminate in very early breast cancer. Both show that these cells colonize distant tissues more efficiently than their later counterparts. [SEE ARTICLE P.552](#) & [LETTER P.588](#)

CYRUS M. GHAJAR & MINA J. BISSELL

The conventional linear model of cancer progression states that the cells of a developing tumour gradually pick up genetic mutations, with cells that accumulate optimal variants eventually acquiring the ability to migrate to and colonize other tissues in the body as metastases. This theory has a huge influence on current views of personalized medicine — for evidence, look no further than the US Cancer Moonshot initiative¹, which recommends extensive sequencing of primary tumours to predict and understand treatment resistance. This represents a reasonable approach only if the disseminated cancer cells (DCCs) that form metastases are derived from cells that populate the primary tumour around the time of its detection. But several lines of evidence^{2–9} indicate that tumour cells can leave the primary site very early during tumour progression and evolve independently at the metastatic site. Online in *Nature*, two papers^{10,11} shed light on the mechanisms of early dissemination for the first time. Strikingly, they offer the first evidence that these early DCCs are more metastasis-competent than cells that leave the tumour at later stages.

The first compelling data to call the linear-progression model into question came from studies of human breast cancer. Genetic analysis of primary breast tumours and corresponding DCCs showed that, at the time of tumour detection, DCCs had fewer genetic alterations than primary cells, implying that DCCs seed the bone marrow early in disease progression, and evolve separately². This theory of parallel progression⁵ was supported by the revelation that 20–30% of patients classified as having ‘non-invasive’ breast cancer have DCCs in their bone marrow^{7,9}. Because up to 8% of ‘non-invasive’ breast cancers recur at distant sites¹², it was assumed that at least some of these early DCCs had metastasis-initiating potential.

Animal models of breast cancer cast further doubt on the linearity of metastasis. In a mouse model in which breast cancer is driven by overexpression of the gene *Her2*, DCCs were detectable in the bone marrow by four weeks of age — just after *Her2* expression begins⁴. Cancerous alterations in the mammary gland are detectable only by electron microscopy at this stage, and palpable primary tumours do not develop for another 14 weeks⁴.

Despite these data, the molecular mechanisms that underlie the early dissemination

of tumour cells have not been identified. To address this, Hosseini *et al.*¹⁰ and Harper *et al.*¹¹ returned to the *Her2*-driven breast-cancer mouse model.

Hosseini *et al.* (page 552) analysed gene expression in epithelial cells that line the ducts of the mammary gland, isolated before the mice reached nine weeks of age, when tumours are not yet palpable. Their analysis suggested that the hormone progesterone drives dissemination from these microscopic early tumours. The authors showed that progesterone triggers secretion of the proteins WNT4 and RANKL from cells expressing the progesterone receptor (PGR), and that these signals imbue epithelial cells that do not express PGR with increased migratory potential (Fig. 1a).

The effect of progesterone becomes less apparent as tumour development progresses. The authors found that both higher cellular density and the increased HER2 protein levels driven by this change induce expression of microRNA molecules that inhibit the gene encoding PGR. In effect, a dissemination-to-proliferation switch occurs when a developing tumour region becomes crowded enough.

Harper *et al.*¹¹ (page 588) identified an invasive cell population in microscopic

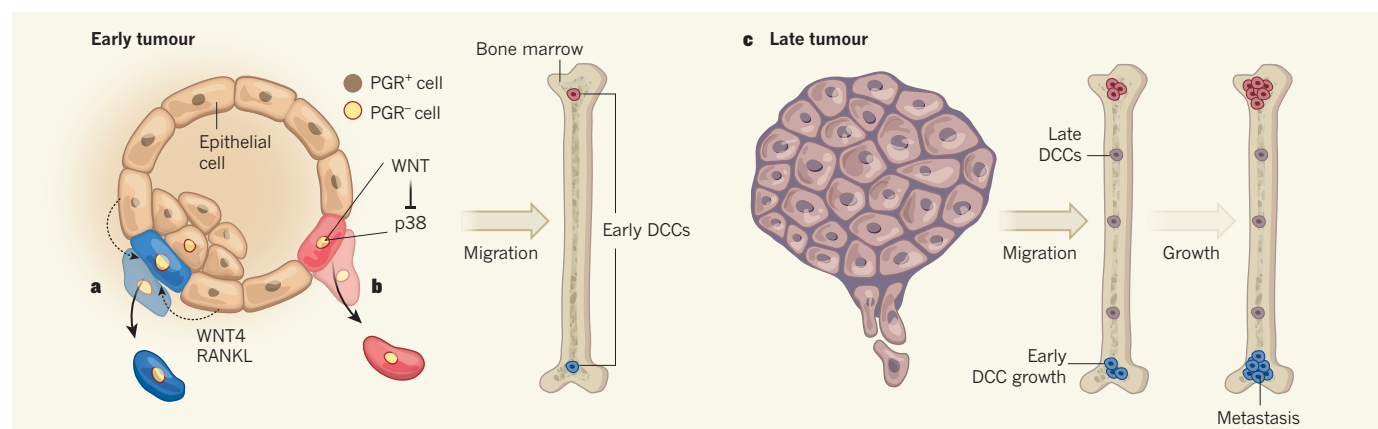


Figure 1 | Mechanisms of early dissemination and metastasis in breast cancer. Two papers outline molecular mechanisms by which cancerous epithelial cells in the mammary glands of mice, driven by overexpression of the gene *Her2*, disseminate from the gland to future sites of metastasis during very early stages of tumour development. **a**, Hosseini *et al.*¹⁰ find that WNT4 and RANKL proteins are secreted from cells in very early tumours that express the progesterone receptor (PGR) protein (PGR⁺ cells). Nearby cells that do not express PGR (PGR⁻) become migratory owing to the

secreted signals and invade the bone marrow as early disseminated cancer cells (DCCs). **b**, Harper *et al.*¹¹ report that increases in the WNT signalling pathway lead to inhibition of the protein p38 and transition of early-tumour cells to an invasive DCC state. **c**, Cells from late-stage tumours also migrate to the bone marrow. Both studies provide evidence that cells that disseminate later in primary-tumour progression form metastases in the bone marrow less efficiently than early DCCs, implying that early DCCs might be a source of metastases.

tumours that is characterized by *Her2*-driven expression of WNT proteins. WNTs counteract the activity of the enzyme p38, which is typically expressed by non-proliferative (dormant) DCCs¹³. Downregulation of p38 leads to a decrease in the epithelial cell–cell junction molecule E-cadherin and to upregulation of several other genes involved in epithelial-to-mesenchymal transition — a change in cell state that facilitates invasion of mammary cells into the bloodstream, and ultimately into other tissues as early DCCs (Fig. 1b). Inhibition of p38 significantly increases the number of circulating tumour cells, as well as the number of DCCs in both the lung and bone marrow, confirming this protein's role in dissemination.

The authors found that DCCs regain some of their epithelial characteristics when they reach the lung and bone marrow. However, it is not clear whether dormancy-associated p38 signalling is restored when the cells enter these sites. Nor is it known whether these early DCCs are less responsive to quiescence-inducing cues from their new microenvironments¹⁴ than cells that leave the primary tumour at later stages.

Are early DCCs less likely to become — or to stay — dormant when they reach distant sites? Perhaps so. Adopting different experimental strategies, both groups demonstrated that early DCCs are inferior to late DCCs in their ability to form primary tumours if implanted in another mammary gland. However, early DCCs are substantially more metastasis-competent, forming metastases faster and more prolifically than their later counterparts (Fig. 1c).

The molecular mechanisms that drive dissemination from early mammary lesions identified in these studies might not apply across all subtypes of breast cancer or to other cancers. Nonetheless, the findings provide a general framework within which to study causality between early DCCs and metastasis — particularly for cancers in which early dissemination is a documented phenomenon, such as skin⁶ and pancreatic⁸ cancers.

These studies have major implications with regard to preventive therapies. First, given the microscopic stage of primary tumour formation at which dissemination of metastasis-competent cells occurs, developing means for early tumour detection may not be sufficient to prevent metastatic disease. And, because cells have probably disseminated by the time a primary tumour is detected, targeting the mechanisms of early dissemination identified in these studies may not be a viable therapeutic strategy either. Thus, we should aim to target the characteristic properties of early DCCs — their long-term survival and therapeutic resistance¹⁴.

As Hosseini *et al.* confirmed from their analysis of human tissue samples, early DCCs differ substantially from the primary tumour at the molecular level. Thus, to learn about

early DCCs, we must increase the frequency with which DCCs are isolated from humans, profiled, and studied functionally in appropriate animal and culture models. This will aid the identification of elusive molecular targets for precision metastasis-prevention therapies based on demonstrable steps in cancer progression, instead of on an assumption of linearity. ■

Cyrus M. Ghajar is in the Public Health Sciences Division's Translational Research Program and the Human Biology Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA.

Mina J. Bissell is in the Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

e-mails: cghajar@fredhutch.org; mjbissell@lbl.gov

SUPRAMOLECULAR CHEMISTRY

Unexplored territory for self-assembly

Cage-like structures can self-assemble from suitable metal ions and organic linkers, but the size of the assemblies was limited. The surprise discovery of a new series of cages opens up fresh horizons for self-assembly. SEE LETTER P.563

FLORIAN BEUERLE

Nature has inspired an efficient way to synthesize large molecular aggregates: self-assembly. In the natural process, multiple copies of subunits (proteins, for example) spontaneously agglomerate into complex, hierarchical architectures such as virus shells (capsids). On page 563, Fujita *et al.*¹ report their use of self-assembly to prepare the largest known, artificially synthesized cage-type object that has a precise atomic composition — an almost spherical shell that assembles from 144 components.

Construction in the macroscopic world is usually associated with building sites, where many builders simultaneously assemble bricks to give shape to an architect's design. But although synthetic chemists can be thought of as molecular architects², construction of molecules at the nanometre scale works quite differently. Conventional synthetic protocols allow target compounds to be constructed only through a linear sequence of steps, and the intermediates formed after each step must be purified in a time-consuming and yield-reducing way. This limits both the size and complexity of target molecules, so that the most complicated structures synthesized so

1. Cancer Moonshot Blue Ribbon Panel Report 2016 (National Cancer Advisory Board, 2016); available at go.nature.com/2gjcac5
2. Schmidt-Kittler, O. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 7737–7742 (2003).
3. Braun, S. *et al.* *N. Engl. J. Med.* **353**, 793–802 (2005).
4. Hüsemann, Y. *et al.* *Cancer Cell* **13**, 58–68 (2008).
5. Klein, C. A. *Nature Rev. Cancer* **9**, 302–312 (2009).
6. Eyles, J. *et al.* *J. Clin. Invest.* **120**, 2030–2039 (2010).
7. Sanger, N. *et al.* *Int. J. Cancer* **129**, 2522–2526 (2011).
8. Rhim, A. D. *et al.* *Cell* **148**, 349–361 (2012).
9. Gruber, I. V. *et al.* *Anticancer Res.* **36**, 2345–2351 (2016).
10. Hosseini, H. *et al.* *Nature* **540**, 552–558 (2016).
11. Harper, K. L. *et al.* *Nature* **540**, 588–592 (2016).
12. Narod, S. A., Iqbal, J., Giannakeas, V., Sopik, V. & Sun, P. *JAMA Oncol.* **1**, 888–896 (2015).
13. Aguirre-Ghiso, J. A., Liu, D., Mignatti, A., Kovalski, K. & Ossowski, L. *Mol. Biol. Cell* **12**, 863–879 (2001).
14. Ghajar, C. M. *Nature Rev. Cancer* **15**, 238–247 (2015).

This article was published online on 14 December 2016.

far consist of no more than several hundred atoms and are only a few nanometres in length (see ref. 3, for example).

Chemists have therefore used self-assembly extensively to make molecular superstructures on different length scales and of diverse shapes and structures. In particular, metallosupramolecular chemistry involves the self-assembly of bifunctional organic ligands such as bipyridyl molecules (which possess two binding sites for metal ions) with ions of metals such as palladium (Pd²⁺). The structures that emerge can be polyhedra, in which the metal ions act as vertices that are connected by edges formed by the organic ligands.

If individual ligands in these compounds can be exchanged easily for other ligands, then the resulting systems can rapidly adjust and rearrange in solution to reach the energetically most stable structure as the sole or major product. Under such dynamic conditions, the structure and topology of the final assembly are mainly governed by three factors: the preferential formation of closed-shell structures that maximize the number of metal-occupied binding sites⁴, rather than polymeric products; the second law of thermodynamics, which maximizes entropy by favouring the formation of many small cages at the expense of larger



50 Years Ago

In the course of two recent geophysical traverses across the Gulf of Guinea, H.M.S. *Hecla* has discovered a number of marked elevations of the ocean bed approximately on a line between St. Helena and the islands of the Bight of Biafra ... The continuous graphical traces made by the precision depth recorder during the traverses indicate that these are rugged topographical features, trending north-east and south-west ... Measured from the abyssal depths from which they rise (2,600 fathoms in the north-west to 3,100 fathoms in the south-east) these features have an elevation exceeding that of the Alps ... The seamounts are near and may form part of the submerged feature, shown on certain American bathymetric charts as "The Guinea Ridge".

From *Nature* 24 December 1966

100 Years Ago

Soon after the outbreak of the war, my father, Lord Roberts, asked the public to lend their glasses for the use of the Army. After two years I think your readers may be glad to have some particulars of the result of his request ... Upwards of 26,000 glasses have been received ... The instruments sent comprise every type, and have been classified and issued according to the needs of different units. Particularly useful have been the fine prismatic glasses sent, which have been allocated to artillery and machine-gun units, according to their power; large mounted telescopes for batteries, deer-stalking telescopes for gunners and snipers, and good old-fashioned non-prismatic racing glasses for detection of the nationality of aircraft, locating snipers signalling by disc, collecting wounded, and musketry instruction.

From *Nature* 21 December 1916

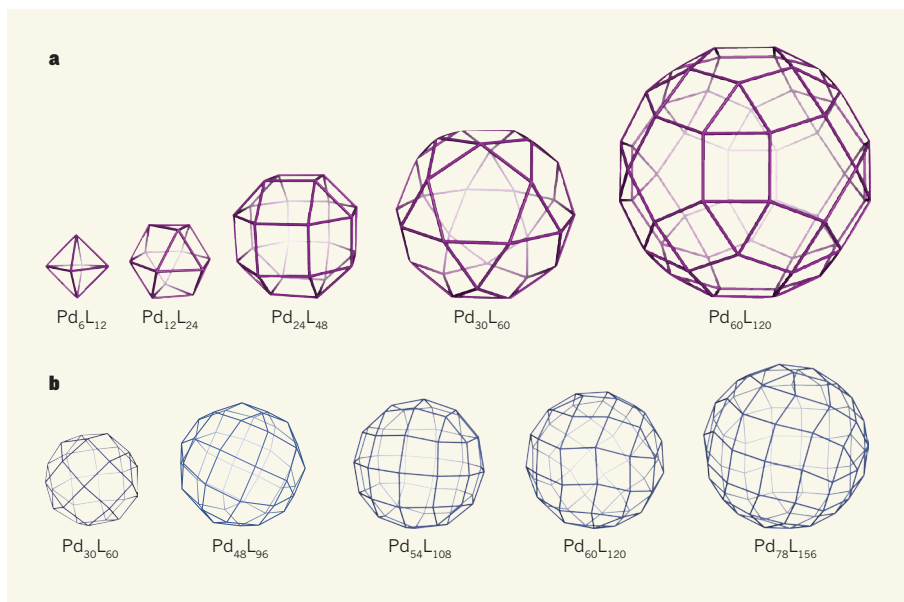


Figure 1 | Self-assembling molecular polyhedra. **a**, Palladium(II) ions (Pd^{2+}) and bipyridyl ligands (L) were predicted⁵ to self-assemble into five possible polyhedra that have the general formula Pd_nL_{2n} , where n can be 6, 12, 24, 30 or 60. The first four of these have been prepared^{6–9} in the past two decades. **b**, While trying to prepare the final member of the series shown in **a**, Fujita *et al.*¹ isolated a structure of formula $\text{Pd}_{30}\text{L}_{60}$, which had an unprecedented topology. They identified this as part of a new infinite series of polyhedra (only five members are shown), and then isolated the next member of this series, $\text{Pd}_{48}\text{L}_{96}$. This achievement opens up research to identify other assemblies of increasing size and complexity.

assemblies; and the preferential formation of 'isotropic' structures that have indistinguishable subcomponents to minimize surface energy and distribute local strain equally throughout the assembly. It therefore follows that the most favourable structures are highly symmetrical objects in the shape of Platonic solids (regular convex polyhedra such as cubes or octahedra) or Archimedean solids (semiregular polyhedra composed of different regular polygons that converge at identical vertices).

Additional design constraints also apply, such as the need for palladium ions to bind to ligands in a square-planar geometry — which implies the convergence of exactly four edges at each vertex. Taken together, the constraints on the self-assembly of palladium ions with bifunctional ligands reduce the number of potential target structures to five cages of formula Pd_nL_{2n} , in which n can be 6, 12, 24, 30 or 60, and L is the ligand⁵ (Fig. 1a).

Over the past two decades, Makoto Fujita's research group has pioneered self-assembly by synthesizing representatives of the first four members of the series: the Pd_6L_{12} octahedron⁶; the $\text{Pd}_{12}\text{L}_{24}$ cuboctahedron⁷; the $\text{Pd}_{24}\text{L}_{48}$ rhombicuboctahedron⁸; and the $\text{Pd}_{30}\text{L}_{60}$ icosidodecahedron⁹. The type of structure that forms depends on the ligand design. For example, subtle changes in the angle formed between the two pyridyl units in a bipyridyl ligand can induce $\text{Pd}_{24}\text{L}_{48}$ stoichiometry to form, rather than $\text{Pd}_{12}\text{L}_{24}$ (ref. 8).

In the present work, Fujita *et al.* targeted the elusive $\text{Pd}_{60}\text{L}_{120}$ rhombicosidodecahedron, the

last representative of the series. However, the authors serendipitously discovered the formation of a $\text{Pd}_{30}\text{L}_{60}$ cage whose single-crystal X-ray structure clearly differed from the previously reported one⁹, and the topology of which did not correspond to any of the Platonic and Archimedean solids. The authors therefore postulated the existence of a new series of polyhedra, which had been reported as theoretical possibilities by mathematicians¹⁰ but never observed in any natural or artificial assemblies: closed-shell frameworks in which eight equally distributed triangles are incorporated into a system of squares (Fig. 1b). These structures are reminiscent of Goldberg polyhedra, which consist of 12 pentagons connected by hexagons, and which are ubiquitous in natural and biological systems — such as fullerene structures and virus capsids.

Impressively, the authors also predicted that the next homologue of the series, $\text{Pd}_{48}\text{L}_{96}$, would be more stable than the isolated compound, and were able to manually separate individual crystals of this larger assembly from the products of their reaction. This cage is by far the most complex molecular structure of precise atomic composition to have been synthesized until now, and is constructed from 144 components through 192 individual metal–ligand interactions.

What is the largest cage structure that could be self-assembled? The extended Goldberg series of polyhedra provides an indefinite number of ever-greater structures, so in principle there is no intrinsic limit to size. However, it will be increasingly difficult to

overcome the entropic penalties associated with the self-assembly of large cages, and to avoid the unwanted but faster formation of smaller cages.

Are there likely to be any practical applications for these giant cages? Investigations into the chemistry and properties of such assemblies might be severely hampered by the difficulties in synthesizing them, especially in bulk quantities rather than just as individual crystals. The structural integrity of the cages, both in solution and in the solid state, is also an unknown crucial issue for any applications. Nevertheless, these huge metal–organic assemblies might encapsulate giant biomolecules such as proteins by forming host–guest interactions, thus stabilizing the biomolecules

and potentially allowing control of their structures in unnatural conditions.

Apart from their value as benchmarks for artificial self-assembly processes, Fujita and co-workers' structures might also inspire interest from other scientific areas. For instance, mathematicians could seek more-exotic topologies as targets for self-assembly, and biologists might search for previously unsuspected topologies in virus capsids or other large biological assemblies. And only time will tell whether Fujita and colleagues' synthetic masterpiece will be the starting point for further journeys into yet-unexplored chemical territory. ■

Florian Beuerle is at the Institute for Organic Chemistry and in the Center for Nanosystems

Chemistry, University of Würzburg, Am Hubland, 97074 Würzburg, Germany. e-mail: florian.beuerle@uni-wuerzburg.de

1. Fujita, D. *et al.* *Nature* **540**, 563–566 (2016).
2. Smulders, M. M. J., Riddell, I. A., Browne, C. & Nitschke, J. R. *Chem. Soc. Rev.* **42**, 1728–1754 (2013).
3. Nicolaou, K. C. *et al.* *Nature* **392**, 264–269 (1998).
4. Kramer, R., Lehn, J. M. & Marquis-Rigault, A. *Proc. Natl Acad. Sci. USA* **90**, 5394–5398 (1993).
5. Harris, K., Fujita, D. & Fujita, M. *Chem. Commun.* **49**, 6703–6712 (2013).
6. Suzuki, K., Tominaga, M., Kawano, M. & Fujita, M. *Chem. Commun.* 1638–1640 (2009).
7. Tominaga, M. *et al.* *Angew. Chem. Int. Edn* **43**, 5621–5625 (2004).
8. Sun, Q.-F. *et al.* *Science* **328**, 1144–1147 (2010).
9. Fujita, D. *et al.* *Chem* **1**, 91–101 (2016).
10. Brinkmann, G. & Deza, M. J. *Chem. Inf. Comput. Sci.* **40**, 530–541 (2000).

CARDIOVASCULAR DISEASE

A turbulent path to plaque formation

Plaque deposits often occur in curved arterial regions with turbulent blood flow. Endothelial cells have been found to respond to blood flow through a previously unidentified signalling pathway that affects plaque build-up. SEE LETTER P.579

VEDANTA MEHTA & ELLIE TZIMA

A key characteristic of the disease atherosclerosis is the gradual accumulation of plaque deposits on the walls of arteries. Plaque is composed of cellular waste, fatty deposits and cholesterol molecules, and is not uniformly distributed in arteries¹. Some plaques can reach a size that obstructs blood flow to organs, causing heart attacks or strokes². On page 579, Wang *et al.*³ propose a mechanism for plaque development that also provides an explanation for plaque-formation patterns.

Blood-flow dynamics have a central role in atherosclerosis development, and the key driving force is shear stress⁴: the frictional force exerted on blood-vessel walls because of blood flow. Shear stress as a result of the uniform laminar blood flow that occurs in straight regions of blood vessels is not considered to be a risk factor for plaque formation⁴. However, curved blood-vessel regions, including those near branch points, have disturbed (turbulent) blood-flow patterns and are more susceptible to plaque development⁴.

How do differences in the mechanical forces exerted on blood vessels result in the promotion or inhibition of plaque formation? Endothelial cells line blood-vessel walls and can sense and distinguish laminar and disturbed blood-flow patterns, which results in changes to endothelial signalling pathways

that ultimately determine whether plaque formation is promoted or inhibited⁴.

YAP and TAZ proteins act as cellular sensors or checkpoints for mechanical forces⁵. These proteins are also master regulators in the Hippo-protein-mediated signalling pathway, which controls organ size

and has a tumour-suppressor function⁶. In atherosclerotic arteries, two of the genes transcribed through the actions of YAP and TAZ are highly expressed^{7,8}; however, direct evidence that links YAP and TAZ to the sensing of mechanical force by endothelial cells and to the development of atherosclerosis has been lacking.

Activation of the YAP and TAZ pathway can be measured by assaying phosphorylation of YAP, movement of the proteins into the nucleus or the expression of their target genes. Using all three assays, Wang *et al.* observed the inhibition of YAP and TAZ activity when endothelial cells grown *in vitro* were subjected to uniform, laminar shear stress. By contrast, YAP and TAZ activity was high when these cells were exposed to disturbed shear stress (Fig. 1).

Wang and colleagues confirmed that YAP and TAZ activity is regulated by blood flow, using an *in vivo* system in which the abdominal

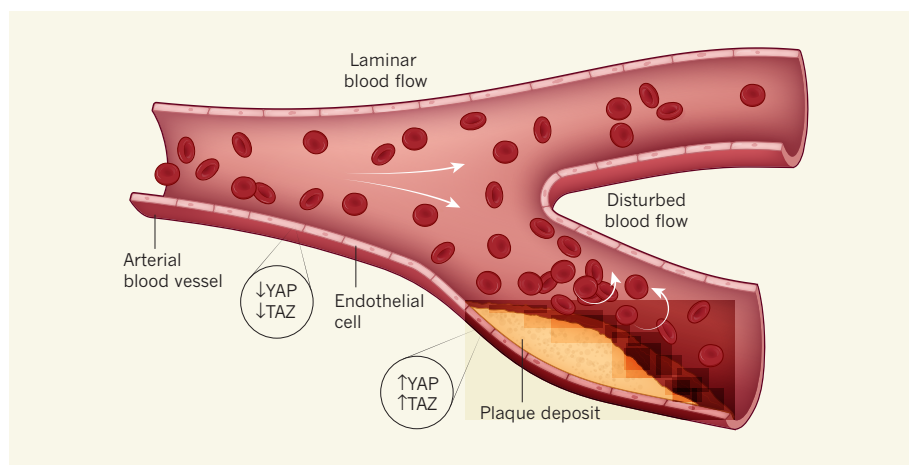


Figure 1 | Endothelial-cell signalling can affect plaque formation. Laminar blood flow parallel to the blood-vessel wall usually occurs in straight regions of arteries. Disturbed, non-laminar blood flow occurs in curved arterial regions, including those near where a vessel branches. The endothelial cells that line blood vessels can sense and respond to these two different types of blood flow⁴. Regions of disturbed blood flow are associated with deposits of plaque, an accumulation of cellular waste and fatty molecules that can obstruct blood flow and potentially cause disease. Wang *et al.*³ report that, in experiments using mouse models and human tissue, endothelial cells adjacent to disturbed blood flow had high YAP and TAZ activity and increased plaque formation. By contrast, endothelial cells adjacent to laminar blood flow have low YAP and TAZ activity and do not have plaque deposits.

artery in rats is clamped. This constriction generates regions of uniform and disturbed shear stress in the same blood vessel⁹. High YAP activity was observed where blood flow was disturbed, and low YAP activity was seen in a region subjected to high uniform shear stress.

To determine whether their findings were relevant to atherosclerosis *in vivo*, Wang and colleagues used a mouse model of atherosclerosis. These animals lack a protein that affects cholesterol metabolism and are susceptible to plaque formation when fed a high-fat diet¹⁰. The mice that developed atherosclerotic plaques had high YAP and TAZ activity in their arteries. The authors also examined samples of human atherosclerotic blood vessels and saw similar high YAP and TAZ activity.

The authors then tested whether manipulating the level of YAP affects plaque formation. Using a version of the atherosclerotic mouse model in which YAP was overexpressed in endothelial cells, they observed that, after four weeks on a high-fat diet, the mice had significantly increased plaque formation as compared with control animals.

To determine the role of YAP in atherosclerosis mediated by disturbed blood flow, the authors subjected atherosclerotic mice¹⁰ to disturbed shear stress through surgery to the carotid artery¹¹. In this system, mice that had also been genetically engineered to have lower endothelial YAP expression had significantly less atherosclerosis than did control animals. Lower atherosclerosis, as compared with control mice, was also seen if the gene expression of TAZ rather than YAP was decreased.

Wang *et al.* propose that the laminar-shear-stress pathway that inhibits YAP and TAZ comprises several molecules that participate in the process of mechanotransduction — the mechanism by which cells convert mechanical signals into biochemical responses. The authors found that laminar shear stress promotes the activation of integrin proteins, promotes the interaction between integrin β_3 and the α_{13} protein and inhibits the protein RhoA, and that these signalling changes subsequently lead to YAP inactivation. Integrin β_3 also has a plaque-promoting role^{12–14}, but how this relates to the plaque-inhibiting role identified by Wang and colleagues is unknown and needs to be investigated.

To explore the plaque-promoting signalling pathways associated with YAP and TAZ activation, Wang and colleagues conducted cellular analyses, including the analysis of messenger RNA sequences. This revealed that YAP and TAZ promote the activation of several inflammatory pathways, including the atherosclerosis-promoting JNK-protein pathway. It is well established that atherosclerosis is a multifactorial disease in which inflammation has a crucial role.

Drugs that lower cholesterol to prevent plaque formation are the most commonly

prescribed medicines in Western countries, and are a first-line therapy for people who have cardiovascular disease. Cholesterol-lowering statin drugs regulate the YAP and TAZ pathway^{15,16}. Whether these drugs protect against plaque formation through modulation of the YAP and TAZ pathway was unknown.

Wang and colleagues treated human cells that express constantly active YAP and TAZ *in vitro* with the statin simvastatin. They found that the treatment did not suppress YAP/TAZ-dependent expression of key genes that promote inflammation and atherosclerosis, indicating that the anti-inflammatory and anti-plaque effects of statins are probably mediated by inhibition of YAP and TAZ activity. This indicates that the YAP and TAZ pathway could be considered as a treatment target for atherosclerosis.

Atherosclerosis is a complex disease in which associated inflammation probably occurs through several different pathways, and this complexity presents an obstacle to successful clinical treatment. Is it possible to target YAP and TAZ specifically in endothelial cells in the arteries? Is suppression of just YAP and TAZ, or the molecules within the YAP and TAZ signalling pathway, sufficient to ameliorate atherosclerosis? Does targeting the pathway also affect the tumour-suppressing function of the Hippo pathway? These questions need to be answered before a therapy to

prevent atherosclerosis can be devised on the basis of YAP and TAZ inhibition. ■

Vedanta Mehta and Ellie Tzima are at the Wellcome Trust Centre for Human Genetics, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 7BN, UK.
e-mail: ellie@well.ox.ac.uk

1. Caro, C. G., Fitz-Gerald, J. M. & Schroter, R. C. *Nature* **223**, 1159–1161 (1969).
2. Ross, R. N. *N. Engl. J. Med.* **340**, 115–126 (1999).
3. Wang, L. *et al. Nature* **540**, 579–582 (2016).
4. Chatzizisis, Y. S., Coskun, A. U., Jonas, M., Edelman, E. R., Feldman, C. L. & Stone, P. H. *J. Am. Coll. Cardiol.* **49**, 2379–2393 (2007).
5. Dupont, S. *et al. Nature* **474**, 179–183 (2011).
6. Piccolo, S., Dupont, S. & Cordenonsi, M. *Physiol. Rev.* **94**, 1287–1312 (2014).
7. Schober, J. M. *et al. Blood* **99**, 4457–4465 (2002).
8. Chaqour, B. & Goppelt-Strube, M. *FEBS J.* **273**, 3639–3649 (2006).
9. Zhou, J. *et al. Proc. Natl Acad. Sci. USA* **109**, 7770–7775 (2012).
10. Zhang, S. H., Reddick, R. L., Burkey, B. & Maeda, N. *J. Clin. Invest.* **94**, 937–945 (1994).
11. Nam, D. *et al. Am. J. Physiol. Heart Circ. Physiol.* **297**, H1535–H1543 (2009).
12. Chen, J., Green, J., Yurdagul, A. Jr, Albert, P., McInnis, M. C. & Orr, A. W. *Am. J. Pathol.* **185**, 2575–2589 (2015).
13. Maile, L. A. *et al. Sci. Transl. Med.* **2**, 18ra11 (2010).
14. Hoshiga, M., Alpers, C. E., Smith, L. L., Giachelli, C. M. & Schwartz, S. M. *Circ. Res.* **77**, 1129–1135 (1995).
15. Sorrentino, G. *et al. Nature Cell Biol.* **16**, 357–366 (2014).
16. Wang, Z. *et al. Proc. Natl Acad. Sci. USA* **111**, E89–E98 (2014).

This article was published online on 7 December 2016.

BIOMATERIALS

Sharks shift their spine into high gear

It emerges that a dogfish shark's spine becomes stiffer as the fish swims faster, enabling the animal to swim efficiently at different speeds. The finding could also provide inspiration for the design of robotic biomaterials.

MATTHEW A. KOLMANN & ADAM P. SUMMERS

The languorous undulation of a shark cruising along a reef gives little hint of the fish's potential to unleash a burst of high-speed movement when pursuing prey. Writing in the *Journal of Experimental Biology*, Porter *et al.*¹ reveal how the structural properties of the non-bony, cartilaginous skeleton of the spiny dogfish shark (*Squalus acanthias*) allow this fish to shift seamlessly between low-speed cruising and high-speed swimming.

A basic principle of aquatic locomotion is that swift swimming requires a stiff spine². A stiffer body decreases drag and increases energy efficiency (Fig. 1). By contrast, acceleration requires a flexible spine to allow a fish to uncurl its body in a sudden rush³. It was

proposed² that thick fibres in a shark's skin increase the stiffness of the fish as it swims faster. This hypothesis is attractive, but has resisted experimental verification because of the difficulty of getting sharks to swim fast in a laboratory while attached to high-tech instrumentation.

Porter *et al.* investigated shark propulsion from a different direction. Rather than considering the fish's exterior, they went right to the core of the matter: the vertebral column, or spine. In bony animals, the vertebrae of the spine are mineralized, rigid, bony structures that do not change shape appreciably during locomotion. Movement of the spine in bony animals occurs through changes in the shape of the intervertebral disks — elastic, but quite firm structures located between individual

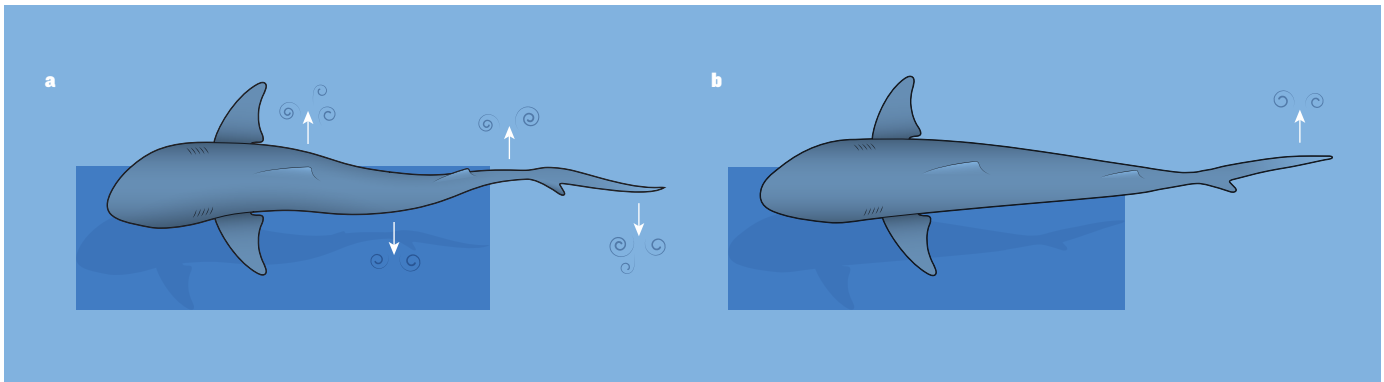


Figure 1 | Shark spine stiffness and swimming efficiency. **a**, Water eddies (blue swirls) around a swimming fish creates drag that slows the animal down. Arrows indicate water moving away from the fish's direction of motion and contributing to drag. If a fish has a flexible spine, this creates an inefficient swimming posture for swift movement because it generates high drag. However, flexible spine movement can be useful for

allowing rapid acceleration. **b**, If a fish has a stiffer spine, it encounters less drag and can swim efficiently at high speed. Porter *et al.*¹ studied the properties of the spine of the dogfish shark (*Squalus acanthias*) at different swimming speeds. The authors found that the fish's spine becomes stiffer as the shark swims faster, thus enabling efficient movement depending on the fish's speed.

vertebrae. These disks consist of an outer ring of fibrous connective tissue encasing a gelatinous core. By contrast, sharks have a mainly cartilaginous skeleton: their vertebrae, as well as their intervertebral disks, are made of cartilage, and the vertebrae are so poorly mineralized that they are potential sites of deformation.

In previous *in vivo* and *in vitro* work by researchers from the same laboratory⁴, sharks had tiny piezo-crystal-based monitors implanted on their vertebrae. These sensors emit and detect ultrasound, allowing inter-crystal distances to be measured with micrometre-scale accuracy. The researchers showed that the vertebrae themselves deformed, not just the intervertebral disks. This earlier work showed that the shark vertebral column could not be simplified as a series of soft connectors between rigid blocks, and instead is a structure that can be variably deformed across its entire length.

To investigate changes to the shark spine during swimming, Porter *et al.* bent excised vertebral columns in the same wave-like motions that are observed during different swimming speeds. The authors used these *in vitro* experiments, together with computer modelling, to determine the length changes that occur across a spinal segment (a series of ten vertebrae and nine intervertebral disks) when the dogfish shark swims. They recorded the corresponding spinal deformations that occurred during movements ranging from the lazy tail wag of a cruising shark to the all-out burst that precedes prey capture. The overall result was a fine-scale understanding of the relationship between spinal deformation and swimming speed.

As a shark swims, the energy used to bend the spine is stored, and is then released when the spine straightens, providing energy for forward motion. Porter and colleagues' work demonstrates that, in swimming sharks, deformation of the vertebrae as well as the

intervertebral disks contributes to the stored energy. The authors also observed that, as sharks swim faster, their spine gets stiffer. The bending of a stiffer spine increases the stored energy that can be used to drive forward motion, and allows the shark to swim faster with greater efficiency. This is an aquatic equivalent of continuously variable transmission, a type of gear-change system found in some motor scooters that is continuously responsive to a wide range of speeds.

What is the structural basis of the shark's remarkable varying spinal properties? Cartilage is a water-laden, fibre-reinforced material. It is a type of in-between substance, neither an elastic solid such as rubber or metal, nor a stirrable fluid such as coffee. Instead, cartilage belongs to the category of viscoelastic materials — materials that resist deformation differently when they change length at different rates. An example of a viscoelastic material is the toy

Silly Putty, which can be drawn hair thin when pulled slowly, but which breaks apart before stretching if it is given a sudden pull.

Intervertebral disks, being made of the gooey composite material cartilage, have the viscoelastic property of becoming stiffer when suddenly strained. This is why they make such poor shock absorbers when someone locks their knees straight to jump off a step. Nevertheless, these structures can undergo gradual compression. Over the course of a day, human intervertebral disks are slowly compressed under the weight they bear, leaving people shorter at night than in the morning⁵. Porter and colleagues' work shows that, like the human version, fish intervertebral disks display the viscoelastic property of an increase

in stiffness when the structure is rapidly strained.

Applications for the discoveries made by Porter *et al.* might arise in the world of soft and bio-inspired robotics, given that microrobots exist that have been designed using the swimming patterns and anatomy of skates, another type of cartilaginous fish, as an inspiration⁶. This discovery of continuously variable transmission in sharks might enable robotic designers to construct light, energy-efficient systems for motion that would require few moving parts and have potentially low levels of wear or little need for replacement parts.

Porter and colleagues' findings highlight the difficulty of characterizing dynamic mechanics in composite biological materials, a research area that could revolutionize design by introducing materials whose physical conformation can shift to fit differing roles. In an age of climate change and increasing environmental pollution, inventors are increasingly looking to nature for inspiration when trying to build clean and efficient machines. What better animals to choose than sharks, given that their capacity for movement has been refined over more than 420 million years^{7,8} of evolution? ■

Matthew A. Kolmann and Adam P. Summers are in the Department of Biology, Friday Harbor Laboratories, University of Washington, Friday Harbor, Washington 98250, USA. e-mails: mkolmann@gmail.com; fishguy@uw.edu

- Porter, M. E., Ewoldt, R. H. & Long, J. H. *J. Exp. Biol.* **219**, 2908–2919 (2016).
- Wainwright, S. A., Vosburgh, F. & Hebrank, J. H. *Science* **202**, 747–749 (1978).
- Blake, R. W. *J. Fish Biol.* **65**, 1193–1222 (2004).
- Porter, M. E. *et al. Zoology* **117**, 19–27 (2014).
- Wasse, J. *Phil. Trans. R. Soc. Lond.* **33**, 87–88 (1724).
- Park, S.-J. *et al. Science* **353**, 158–162 (2016).
- Wilga, C. D. & Lauder, G. V. *Nature* **430**, 850–850 (2004).
- Maisey, J. G. *J. Fish Biol.* **80**, 918–951 (2012).

This article was published online on 14 December 2016.

PLANETARY SCIENCE

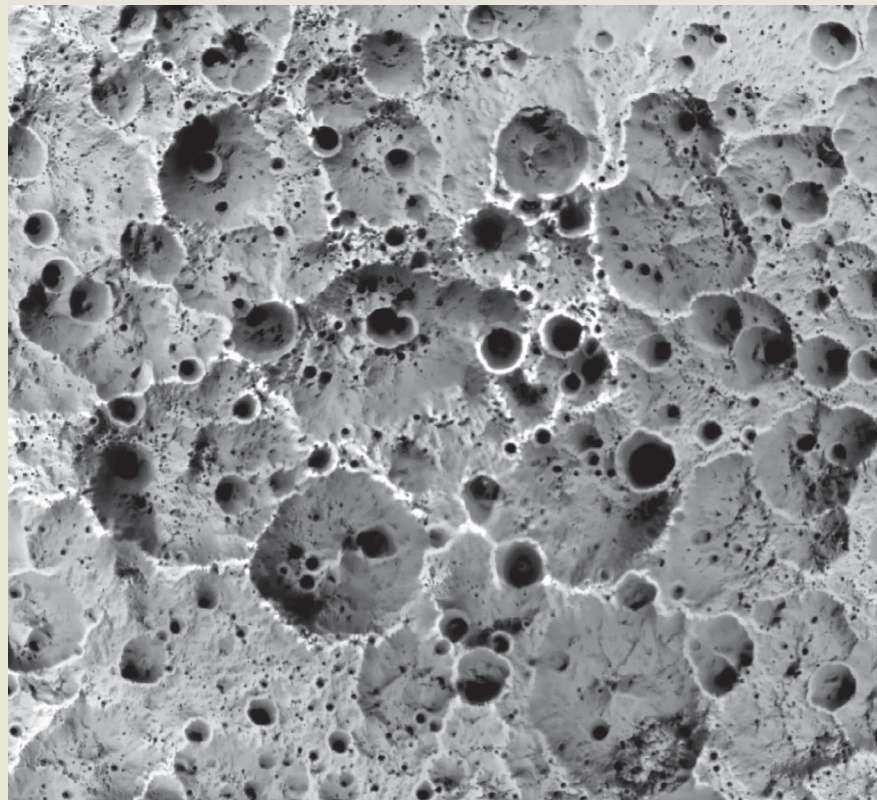
Frozen in darkness

In 2014, water vapour was detected around Ceres, a dwarf planet in the asteroid belt. NASA's Dawn spacecraft, in orbit around Ceres since March 2015, subsequently found water ice on the dwarf planet, in a small, mid-latitude crater named Oxo. Now, writing in *Nature Astronomy*, Platz *et al.* show that polar regions of Ceres that lie in perpetual darkness trap and preserve water ice, providing valuable information about the processes that create water reservoirs on planetary bodies (T. Platz *et al. Nature Astron.* <http://dx.doi.org/10.1038/s41550-016-0007>; 2016).

Permanently shadowed areas on airless planetary bodies, usually found in craters at polar latitudes, have favourable conditions for retaining water ice. Without sunshine, these regions are extremely cold (a few tens of kelvin), and can therefore efficiently trap water molecules and retain them for geological timescales. This cold-trapping process is known to be active on the Moon and on Mercury.

Platz and colleagues analysed images obtained by Dawn's Framing Camera to reconstruct complete maps of the shadowed areas of the northern polar region of Ceres (pictured). They find more than 600 craters in perennial shadow, 10 of which exhibit bright features. One of these features has a (partially) illuminated portion, allowing Platz *et al.* to perform a spectral analysis using Dawn's imaging spectrometer. The spectra show clear signatures of water ice, providing definitive evidence for the nature of these bright deposits.

The authors also constrain the timescale



of water-ice production on Ceres using the fact that ice deposits evolve through two competing mechanisms: slow accumulation by cold-trapping and destruction by impact 'gardening', the overturning of planetary soil by impacts. The authors estimate that the deposits are young (not more than a few hundred thousand years old), which implies

that ice delivery is continuously active. In addition, the deposits are relatively scarce (compared with, for example, those on Mercury), suggesting that Ceres is unable to retain much water or didn't have much to begin with. [Luca Maltagliati](#)

This article was published online on 19 December 2016.

CONDENSED-MATTER PHYSICS

Quantum mechanics in a spin

Quantum spin liquids are exotic states of matter first predicted more than 40 years ago. An inorganic material has properties consistent with these predictions, revealing details about the nature of quantum matter. [SEE LETTER P.559](#)

LEON BALENTS

The phenomenon of magnetism, discovered thousands of years ago, arises from the alignment of electron magnetic moments known as spins. But if these spins do not align, they can form a

truly quantum state called a quantum spin liquid (QSL). On page 559, Shen *et al.*¹ report measurements of exotic spin excitations in an inorganic material (of ytterbium, magnesium, gallium and oxygen; YbMgGaO₄). The authors' observations suggest that YbMgGaO₄ forms a QSL that is closely analogous to a state

of matter associated with electrons in a metal. The appearance of electron-like particles in such a material is surprising, and indicates extraordinary 'quantum entanglement' of the underlying spins.

The concept of a QSL was first introduced² in 1973 by the physicist Philip Anderson, who described such a system as "resonating", indicating the presence of quantum superposition — in quantum mechanics, reality is represented by a wavefunction in which physical states of a system can be added together like numbers in arithmetic. An extreme example of quantum superposition is Schrödinger's famous cat, whose wavefunction is the sum of the state of the living cat and that of the dead cat. The QSL is a close relative of Schrödinger's cat that incorporates long-range entanglement (a superposition involving many widely separated spins). According to modern quantum theory, this type of entanglement is so stable

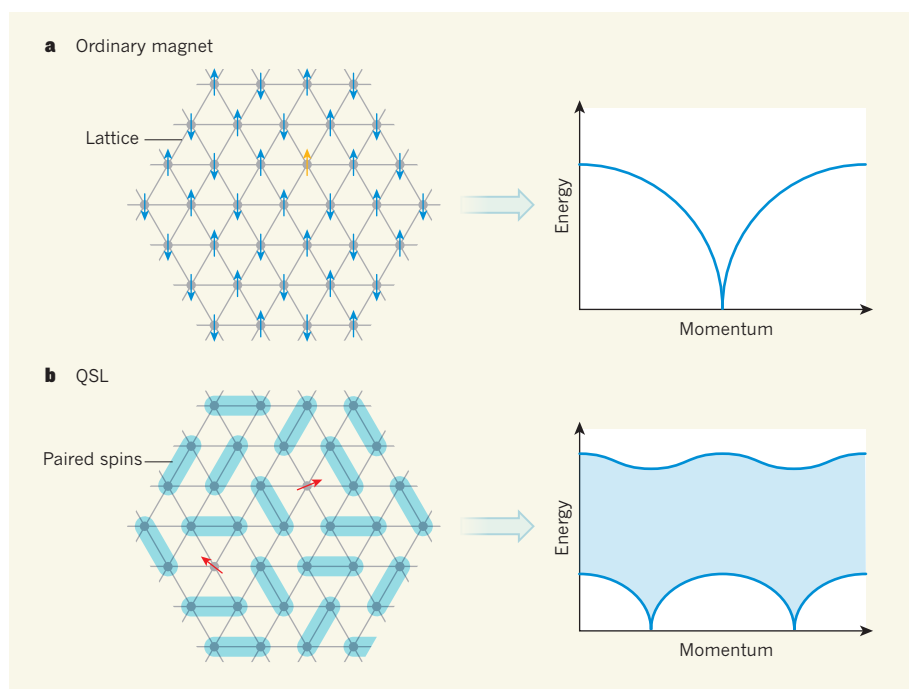


Figure 1 | Identifying a quantum spin liquid. **a**, In a neutron-scattering experiment, an ordinary magnet can experience an excitation called a magnon (orange arrow) — a spin that is flipped with respect to the ordered pattern of the surrounding spins (blue arrows) in the lattice. For a fixed value of the momentum, such an excitation has a specific energy. **b**, Conversely, in a quantum spin liquid (QSL), neutron scattering can create two ‘spinons’ (red arrows); the surrounding spins are paired in ‘valence’ bonds. The two spinons give rise to a continuous energy spectrum: for a fixed momentum, the energy can take a wide range of values. Shen *et al.*¹ report the observation of such a spectrum in the inorganic compound YbMgGaO₄, suggesting that this material could be a QSL.

that a QSL constitutes a new phase of matter at zero kelvin^{3,4}.

The pattern of entanglement in a QSL state can be disrupted locally to form objects called quasiparticles that behave like ordinary particles. Such objects can have substantially different properties from those of the underlying microscopic spins. A dramatic example is the appearance of electron-like particles called fermions in a QSL — these particles obey Pauli’s exclusion principle (they cannot share a single quantum state), whereas spins do not. Such behaviour enables the formation of a spinon metal, a material that is a conductor of spin and heat, but an electrical insulator.

A spinon metal has been implicated in the organic crystals κ -(BEDT-TTF)₂Cu₂(CN)₃ and EtMe₃Sb[Pd(dmit)₂]₂, which have been the most-studied QSLs experimentally for more than a decade⁵. Here, the presence of a spinon metal was deduced indirectly from thermodynamic and thermal-transport measurements. The spins in these crystals are arranged in an approximately triangular lattice — such a geometry prevents a simple alignment of spins that could disrupt QSL formation.

Shen and colleagues study the inorganic material YbMgGaO₄, which, like the organic crystals, contains a triangular lattice of spins — in this case, Yb³⁺ ions separated by non-magnetic layers⁶ (see Figure 1 of the paper¹). However, YbMgGaO₄ is atomically dense

(there is one spin per 5 square ångströms in each layer⁷; for comparison, the area per spin in κ -(BEDT-TTF)₂Cu₂(CN)₃ is more than 100 Å²; ref. 8), and each spin carries a roughly 50% larger magnetic moment than its organic counterpart⁶. These distinctions, and the absence of hydrogen, make YbMgGaO₄ a perfect candidate for neutron scattering, the gold standard of magnetic measurements.

“YbMgGaO₄ is a perfect candidate for neutron scattering, the gold standard of magnetic measurements.”

In inelastic neutron scattering, a neutron transfers energy and momentum to a material, which manifests as an excitation of the electron spins. Using many neutrons, one can determine the number of such excitations that have a given energy and momentum. The form of this spectral density is very different in an ordinary magnet and a QSL (Fig. 1). In a magnet, the excitation, called a magnon, is simply a spin that is flipped with respect to the ordered pattern of the surrounding spins. Such an excitation behaves like a particle and therefore has a definite relationship between its energy and its momentum. As a result, the spectral density at fixed momentum is non-zero only at specific energies. By contrast, in a QSL, the excitation is two ‘spinons’, which share

the energy and momentum that are imparted to the material. Consequently, the spectral density is non-zero for a large range of energies — a continuum.

Shen and collaborators report the observation of such a continuum in YbMgGaO₄. The spins in this material have been shown⁶ to remain unaligned at temperatures as low as 30 mK — strongly suggesting that this feature persists to 0 K — and thermodynamic quantities such as the specific heat and magnetic susceptibility have features quite similar to those of the organic crystals. The authors’ data reveal a much more detailed view of this QSL state, showing a continuum spectrum that is devoid of any sharp excitation peaks; it has maxima in energy and an overall intensity that are momentum dependent. Shen *et al.* compare their material’s spectral density at 70 mK with theoretical predictions for various QSL states, finding that the best agreement is with a spinon metal.

The authors determine the nature of YbMgGaO₄ purely empirically — its QSL behaviour now needs to be understood theoretically. Microscopically, the forces between Yb³⁺ spins are direction dependent and somewhat complicated, and their effects have not been definitively determined. Moreover, the non-magnetic layers between the triangular planes in YbMgGaO₄ have a random arrangement of Mg and Ga atoms — this must induce some disorder in the spins, but its role in QSL formation is unknown.

If the authors’ interpretation is correct, their results constitute the first momentum- and energy-dependent spectroscopy of a spinon metal. Note that a simultaneously published study⁹ has reported consistent experimental results for YbMgGaO₄, and that a possible spinon continuum has been observed in the mineral herbertsmithite¹⁰, which has been suggested to be a different type of QSL. Further exploration of spinon metals could reveal rich physics — for example, spinons are predicted to interact strongly through ‘emergent gauge fields’, whose effects on the spectral density are unknown. It is exciting to imagine these beautiful theoretical ideas realized in the laboratory. ■

Leon Balents is at the Kavli Institute for Theoretical Physics, University of California, Santa Barbara, California 93106, USA.
e-mail: balents@kitp.ucsb.edu

- Shen, Y. *et al.* *Nature* **540**, 559–562 (2016).
- Anderson, P. W. *Mater. Res. Bull.* **8**, 153–160 (1973).
- Balents, L. *Nature* **464**, 199–208 (2010).
- Savary, L. & Balents, L. *Rep. Prog. Phys.* **80**, 016502 (2017).
- Kanoda, K. & Kato, R. *Annu. Rev. Condens. Matter Phys.* **2**, 167–188 (2011).
- Li, Y. *et al.* *Sci. Rep.* **5**, 16419 (2015).
- Li, Y. *et al.* *Phys. Rev. Lett.* **115**, 167203 (2015).
- Geiser, U. *et al.* *Inorg. Chem.* **30**, 2586–2588 (1991).
- Paddison, J. A. M. *et al.* *Nature Phys.* <http://dx.doi.org/10.1038/nphys3971> (2016).
- Han, T.-H. *et al.* *Nature* **492**, 406–410 (2012).

2016

EDITORS' CHOICE

Extracts from selected News & Views articles published this year.

CRYSTALLOGRAPHY

RESOLUTION BEYOND THE DIFFRACTION LIMIT

Jian-Ren Shen (*Nature* 530, 168–169; 2016)

The resolution at which structures can be visualized using X-ray crystallography depends on the diffraction limit, which in turn depends on the regular arrangement of molecules within crystals. The degree of regularity is often low in crystals of biomacromolecules, creating a major barrier to visualizing these molecules at the atomic level. Ayer *et al.* report an approach to solving molecular structures at resolutions higher than the diffraction limit. They propose that molecules in crystals can be considered as rigid units, and that disorder is caused by translational displacements of these units from their lattice positions. These displacements cause scattered X-rays to produce a vague pattern of light and shadowy regions, which contains information about the molecules' structure at resolutions beyond the limit of normal Bragg diffraction. The authors used this information to extend the resolution of a membrane-protein complex from 4.5 Å to 3.5 Å. The technique is potentially a great step forward for those seeking high-resolution structural information for many 'poorly diffracting' protein crystals.

Original research: *Nature* 530, 202–206 (2016).

ATOMIC PHYSICS

A MILESTONE IN QUANTUM COMPUTING

Stephen D. Bartlett (*Nature* 536, 35–36; 2016)

Quantum-savvy entrepreneurs are already bringing the first quantum computer processors out of the physics laboratory and onto the market. But these devices are mostly designed to perform just one function and cannot be programmed to run different algorithms. It would therefore be advantageous to build a fully fledged quantum computer that could be programmed to run anything we might want. In particular, it might execute the complex quantum algorithms that researchers think will solve today's intractable problems in quantum chemistry, materials science and data security. Debnath *et al.* present a small but fully programmable quantum computer consisting of five quantum bits (qubits), and they demonstrate its functionality by running several simple quantum algorithms. In all of these demonstrations, the resulting error rate is consistent with the authors' observations of how their qubits work in isolation, showing that the qubits can be used together in more-sophisticated algorithms in the future. The next challenge for such technologies is to demonstrate that quantum error correction can bring error rates down to negligible levels.

Original research: *Nature* 536, 63–66 (2016).



ZACHARY HUANG

EVOLUTION

INSECT INVASIONS AND NATURAL SELECTION

Amro Zayed (*Nature* 539, 500–502; 2016)

The success of social-insect invaders is paradoxical, because they have a sex-determination system that gives rise to many sterile or inviable males in small founding populations. Gloag *et al.* show how, during the early stages of an invasion of the Asian honeybee *Apis cerana*, the action of natural selection can lessen this problem. Sex determination in bees is governed by a gene called *csd*. Different versions of genes are called alleles. Under conditions in which fertilized eggs have two identical alleles of *csd*, fertilized eggs produce males that are sterile or inviable. The authors speculated that a form of natural selection, called balancing selection, would have a role in reducing imbalances in the frequency of *csd* alleles. With balancing selection, individuals that have rare *csd* alleles would be expected to have high fitness. Gloag *et al.* sequenced the *csd* gene in the *A. cerana* population to determine how it evolved during the invasion. They found that the frequency of the *csd* alleles started to converge on the frequency expected if balancing selection is assumed.

Original research: *Nature Ecol. Evol.* <http://dx.doi.org/10.1038/S41559-016-0011> (2016).

GEOLOGY

EVIDENCE OF LIFE IN EARTH'S OLDEST ROCKS

Abigail C. Allwood (*Nature* 537, 500–501; 2016)

When did life first arise on Earth? Nutman *et al.* analysed 3.7-billion-year-old rocks in the Isua Greenstone belt in Greenland. Within the rocks can be seen ancient ripple marks and piles of rock fragments deposited during an ancient storm. In the middle of it all are structures resembling stromatolites: layered structures that form through microbially influenced accretion of sediment. If these are really the figurative tombstones of our earliest ancestors, the implications are staggering. Earth's surface 3.7 billion years ago was a tumultuous place. If life could find a foothold here, and leave such an imprint that vestiges exist, even though only a minuscule sliver of rock is all that remains from that time, then life is not a fussy, reluctant and unlikely thing. Give life half an opportunity and it'll run with it.

Original research: *Nature* 537, 535–538 (2016).

366 DAYS:
the year in scienceROBOTICS
GENERATION SOFTBarbara Mazzolai &
Virgilio Mattoli
(*Nature* 536, 400–401; 2016)

Robots are typically used in manufacturing contexts that involve well-structured environments. But if these machines were moved into 'real' environments, they would have to cope with uncertain situations and adapt to changing conditions — tough problems to solve using conventional technology made from hard materials. Robots made from soft, deformable materials would be better able to grasp and manipulate unknown objects, and to move on unstructured and rough terrains. Wehner *et al.* present the first robot that completely lacks rigid structures and control systems. The octopus-shaped robot has eight arms moved by a mechanism that relies on the expansion of embedded, inflatable components integrated into a fluidic-pneumatic network, powered by a liquid fuel. Although soft robotics is still in its infancy, it holds great promise for applications such as servicing and inspecting machinery, search-and-rescue operations, and exploration.

Original research: *Nature* **536**, 451–455 (2016).

GRAVITATIONAL WAVES
DAWN OF A NEW ASTRONOMYM. Coleman Miller (*Nature* 531, 40–42; 2016)

Albert Einstein discovered that binary stars and other sources should generate gravitational waves. Unfortunately, he also found that any imaginable source would produce waves so weak that detection was inconceivable using the technology of the day. But this inconceivable detection has now been reported by Abbott *et al.* (the LIGO Scientific Collaboration and the Virgo Collaboration) in *Physical Review Letters*. The authors describe the detection of the signal GW150914 from gravitational waves generated by the merger of two black holes. Astronomers previously had only three types of messenger from space beyond our Solar System: photons, neutrinos and high-energy cosmic rays. Gravitational waves can now be added to this short list. Opening this window will reveal astronomical events that had only been hypothesized. The signal has also provided the most direct confirmation yet of the existence of event horizons — the boundaries beyond which nothing can escape a black hole's gravitational field.

Original research: *Phys. Rev. Lett.* **116**, 061102 (2016).

GENOMICS
FROM SEA TO SEASusan L. Williams (*Nature* 530, 290–291; 2016)

Eelgrass (*Zostera marina*) is an unlikely model for plant evolution, but is a useful one because it has undergone major habitat shifts: it evolved from marine algae into a terrestrial flowering plant, then moved back to the sea again. Olsen *et al.* describe the complete genome sequence of eelgrass. The sequence reveals that, in moving from calm lakes and ponds to the rough, salty ocean, eelgrass lost several key gene groups. For evolutionary biologists, the genome represents a missing piece in the puzzle of angiosperm evolution. For marine ecologists, the genome is a powerful tool for uncovering the adaptations that allow the plant to thrive in a wide range of environmental conditions. This ability to adapt might be the key to surviving environmental changes such as ocean acidification, warming and freshening that are occurring under global climate change.

Original research: *Nature* **530**, 331–335 (2016).

FORUM: Neuroscience
VIRTUAL REALITY EXPLORED
(*Nature* 533, 324–325; 2016)

Neuroscientists are increasingly using virtual reality (VR) to facilitate studies of animal behaviour, but whether behaviour in the virtual world mimics that in real life is a matter for debate.

THE BEST OF BOTH WORLDS

Matthias Minderer & Christopher D. Harvey

VR allows researchers to define explicitly and exhaustively the sensory cues that carry information about the virtual world. It offers the means to add or remove sensory cues to test the contribution of each one to a neural code, and to build up a 'minimal' set of stimuli needed to produce a given behaviour or neural activity pattern. A second benefit comes from the ability to redefine the laws that link the subject's actions to changes in its world. Third, VR increases the range of tools available to measure neural activity.

A WORLD AWAY FROM REALITY

Flavio Donato & Edvard I. Moser

Pressing concerns are raised when VR is used to study higher-order computations. Navigation reflects the integration of many sensory inputs. But in VR, these elements are often not coordinated, and the animal must overcome discrepancies between visual cues that follow movements and cues that are static in VR, such as smell. Such discrepancies might alter the activity of space-encoding neurons to reflect only information coordinated to motion, such as visually changing landmarks and accumulated distance, at the expense of other cues. This could lead researchers to overestimate the contribution of visual inputs to navigation.



Redefining the invertebrate RNA virosphere

Mang Shi^{1,2*}, Xian-Dan Lin^{3*}, Jun-Hua Tian^{4*}, Liang-Jun Chen^{1*}, Xiao Chen^{5*}, Ci-Xiu Li^{1*}, Xin-Cheng Qin¹, Jun Li⁶, Jian-Ping Cao⁷, John-Sebastian Eden², Jan Buchmann², Wen Wang¹, Jianguo Xu¹, Edward C. Holmes^{1,2} & Yong-Zhen Zhang¹

Current knowledge of RNA virus biodiversity is both biased and fragmentary, reflecting a focus on culturable or disease-causing agents. Here we profile the transcriptomes of over 220 invertebrate species sampled across nine animal phyla and report the discovery of 1,445 RNA viruses, including some that are sufficiently divergent to comprise new families. The identified viruses fill major gaps in the RNA virus phylogeny and reveal an evolutionary history that is characterized by both host switching and co-divergence. The invertebrate virome also reveals remarkable genomic flexibility that includes frequent recombination, lateral gene transfer among viruses and hosts, gene gain and loss, and complex genomic rearrangements. Together, these data present a view of the RNA virosphere that is more phylogenetically and genomically diverse than that depicted in current classification schemes and provide a more solid foundation for studies in virus ecology and evolution.

RNA viruses are likely to exist in every species of cellular life¹. Despite this ubiquity, much of our knowledge of the biodiversity and evolution of RNA viruses, as well as their range of genomic structures, comes from those viruses that can be cultured and that act as agents of disease in humans or economically important animals and plants. However, these only represent a tiny fraction of eukaryotic diversity. This sparse sampling is apparent from studies of invertebrate viruses. Although invertebrates comprise the vast majority of the Metazoa (animals), little is known about the nature of the ‘virosphere’ of these organisms². Metagenomic studies of invertebrate viruses have only recently been undertaken but often reveal far greater viral biodiversity than seen in vertebrates^{3–8}. Arthropods, for example, commonly act as viral vectors and studies of arthropod RNA viruses have revealed that changes in genome size, structure and segmentation have occurred more frequently and on a larger scale than previously realized^{3,9}, with some arthropod viruses likely to be ancestors of those that infect vertebrates³. However, these studies are of limited scope and there are still substantial gaps in our knowledge of RNA virus biodiversity at both the phylogenetic and genomic scales for most invertebrates, a fact that may have important implications for our understanding of virus evolution, ecology and emergence¹⁰. We describe here a large-scale meta-transcriptomic survey of diverse invertebrate taxa aimed at revealing the hidden diversity of RNA viruses. The data obtained enable us to re-examine and re-define the invertebrate virosphere, providing a new perspective on the fundamental patterns and processes of viral evolution.

RNA viruses in invertebrates

We performed deep transcriptome sequencing on more than 220 invertebrate species, representing 9 metazoan phyla (Arthropoda, Annelida, Sipuncula, Mollusca, Nematoda, Platyhelminthes, Cnidaria, Echinodermata, and the Chordata subphylum Tunicata), most of which have not previously been screened for viruses (Supplementary Table 1). Accordingly, we extracted total RNA from these species and prepared 87 RNA sequencing (RNA-seq) libraries for Illumina HiSeq sequencing

(Supplementary Table 1). In total, we generated 6 trillion bases of 90–100 bp paired-end reads that were assembled *de novo* for virus characterization.

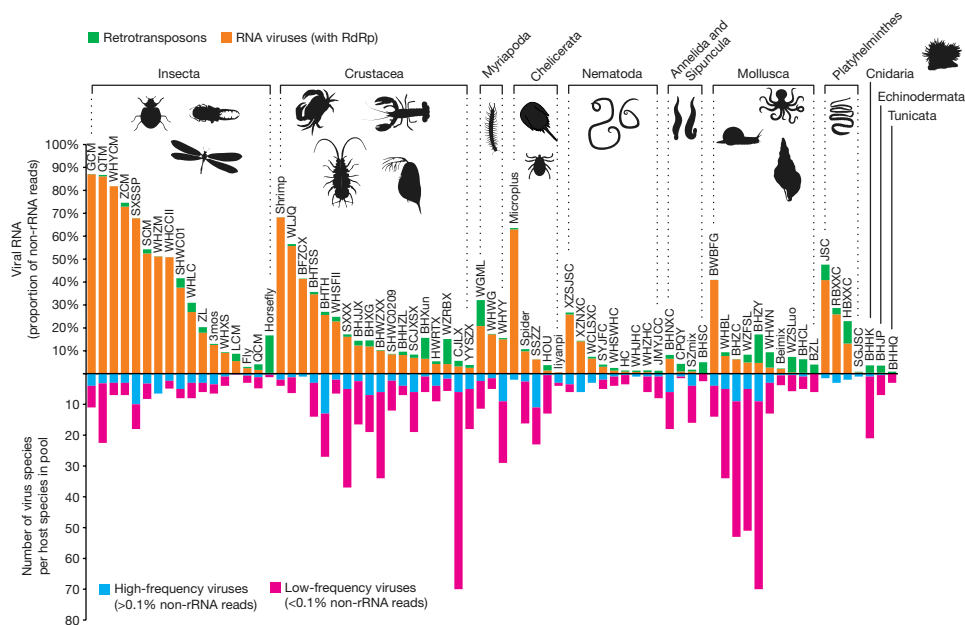
These transcriptome data allowed us to identify at least 1,445 phylogenetically distinct virus genomes or genome segments that contained an RNA-dependent RNA polymerase (RdRp) domain (Supplementary Table 2). The majority of these virus genomes have greater than 20-fold coverage and are sequenced to their complete or near-complete length. Sequence alignments and structural comparisons revealed extensive sequence divergence within these newly discovered RdRp domains, with most sharing less than 40% amino acid identity with those RNA viruses described previously.

To assess the amount of viral RNA in each library, we removed all rRNA reads, including those from the host species, and determined the proportion of the remaining sequence data that mapped to viral RNA. This revealed that viral RNA comprised from 0.05% to 87% of the total RNA sequenced (rRNA excluded) within each library, although the very high levels in some cases may reflect degradation or inefficient extraction of host RNA (Fig. 1). Each library contains 1–20 virus species per host species and 1–6 virus species that represent more than 0.1% of the total RNA sequenced (rRNA excluded) (Fig. 1). Although some libraries contain far higher numbers of virus transcripts, most have low levels of RNA and may therefore be associated with other cellular organisms that are present within the host (see below).

These transcriptome data also contain a substantial proportion of transcripts that carry divergent reverse transcriptase enzymes, potentially derived from retrotransposons (Fig. 1). These can be distinguished from RNA viruses by their replicase components, the lack of consistently inherited structural proteins and the presence of DNA copies. Additionally, although RNAs produced by DNA viruses (including bacteriophages) and bacteria were present in the transcriptome data, these were generally at lower quantities than RNA viruses and will not be discussed further.

¹State Key Laboratory for Infectious Disease Prevention and Control, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Changping, 100206 Beijing, China. ²Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, the University of Sydney, Sydney, New South Wales 2006, Australia. ³Wenzhou Center for Disease Control and Prevention, Wenzhou, 325001 Zhejiang, China. ⁴Wuhan Center for Disease Control and Prevention, Wuhan, 430015 Hubei, China. ⁵Guangxi Mangrove Research Center, Beihai, 536000 Guangxi, China. ⁶Systems Biology and Bioinformatics Group, School of Biological Sciences, Faculty of Sciences, University of Hong Kong, Hong Kong, China. ⁷National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention, Shanghai, China.

*These authors contributed equally to this work.



The top graph shows the percentage of non-rRNA reads mapped to viral RNA (RdRp; orange bars) and retrotransposons (green bars) in each library. The short name of each library is shown on top of each bar, while major host classifications are shown above the bar graph. The bottom graph shows a summary of the normalized number of virus species within each library. The total number is further subdivided to identify those RNA viruses at either high (blue) or low (magenta) frequency.

Evolution of invertebrate RNA viruses

To place these newly discovered viruses within the context of known viral biodiversity, we collated NCBI reference virus genomes from all established families and floating genera of RNA viruses, as well as non-reference virus genomes from unclassified taxa. The RdRp is the only conserved-sequence domain across all RNA viruses and was therefore used for phylogenetic inference. Phylogenetic analysis revealed that the genetic diversity of the newly discovered RNA viruses surpassed that described previously and could not always be easily incorporated into current virus classifications (Fig. 2; see Supplementary Data 1–21 for detailed phylogenies). In particular, many of the newly discovered viruses occupy topological positions that fall between families or genera and thus fill major phylogenetic gaps, so that RNA viruses now occupy a more continual spectrum of phylogenetic diversity.

To describe and accommodate the extraordinary diversity of viruses discovered here better, we merged previously defined virus families, orders and floating genera, to produce 16 clades of RNA viruses. For simplicity, we have abbreviated these clades as 'Astro', 'Birna', 'Hepe-Virga', 'Hypo', 'Luteo-Sobemo', 'Narna-Levi', 'Bunya-Arena', 'Mono-Chu', 'Orthomyxo', 'Nido', 'Partiti-Picobirna', 'Permutotetra', 'Picorna-Calici', 'Reo', 'Tombus-Noda' and 'Toti-Chryso', reflecting the presence of representative viral families or orders within each clade (Fig. 2 and Supplementary Data 1–20). Notably, these clades resemble, but do not necessarily correspond to, the 'supergroups' of RNA viruses proposed previously¹¹. We also identified at least five clades of RNA viruses in which RdRp domains are so divergent that they might be considered as new virus families or orders, although phylogenetic analyses of such divergent taxa should be treated with caution (Supplementary Data 21). Reflecting the location of their sampling, we provisionally named these divergent lineages after ancient Chinese states from the Chunqiu period, specifically; 'Yuevirus', 'Qinvirus', 'Zhaovirus', 'Weivirus' and 'Yanvirus'.

Since our sample processing involves the entire individual invertebrate, it is possible that a substantial proportion of the viruses discovered here were associated with undigested food, gut microflora or parasites that exist within the organisms investigated. We therefore estimated the proportion of each viral transcript within the library and assumed that the more common the virus, the more likely that it was associated with that host (although this may not equate to active infection). Generally, those viruses that made up a higher proportion of total RNA levels ($>0.1\%$ total RNA) were not closely related to those known to infect vertebrates, plants or fungi, suggesting that viral RNA quantity may be a useful indicator of their true host. To assess the likely host species further, we screened for endogenous virus elements (EVEs)

related to the exogenous viruses described here¹². Although some viruses, such as the *Picornavirales*, rarely possess EVEs, and there is little genome data for species within the Annelida and Mollusca phyla, the EVE data helped confirm the host taxon by establishing their evolutionary ancestry in that host. In particular, the host taxa containing EVEs often closely matched those containing related exogenous viruses (Supplementary Data 3, 7–9, 11, 18, 20, 21). However, the EVE data also suggested alternative or additional hosts in a number of cases (Supplementary Data 5, 6, 21). For example, the highly divergent new RNA virus (Weivirus) identified in the mollusc transcriptome was related to EVEs identified in alveolates (protist) genomes (Supplementary Data 21). Finally, we also examined the presence of variant genetic codes as a guide to the likely host organisms. The most common variant genetic codes observed were the invertebrate mitochondrial code (Supplementary Data 6, 11) and the ciliate, dasycladacean and hexamita nuclear codes. The latter is found in the new RNA virus Zhaovirus (Supplementary Data 21), as well as in a cluster of viruses from the Tombus–Noda clade (Supplementary Data 19), indicating that these viruses are more likely to be associated with protists than with invertebrates.

Overall, the host spectrum for the RNA viruses described here is broad, including different phyla and sometimes different kingdoms (Supplementary Data 1–21). Much of our sampling was directed towards the Arthropoda, meaning that definite statements on host range cannot be made. Despite this bias, the diversity of arthropod viruses is notable as they appear in multiple lineages within each major clade (Extended Data Fig. 1 and Supplementary Data 1–21). Also of note were the phyla Mollusca, Annelida, and Sipuncula (collectively the superphylum Lophotrochozoa) that diverged early from Nematoda and Arthropoda in the metazoan phylogeny¹³. Notably, the viromes of these phyla either contained extremely divergent viruses (such as in the Bunya–Arena and Orthomyxo clades; Supplementary Data 7 and 9, respectively) or had substantial overlap with the arthropod virome (for example, several viruses in the aquatic picorna-like clade are commonly present in both Crustacea and Lophotrochozoa; Supplementary Data 14). Although only a limited number of species were available for the remaining phyla, it is notable that the Platyhelminthes and Cnidaria did not contain particularly divergent viruses, despite their basal position within Metazoa. Finally, although the phylum Echinodermata and the subphylum Tunicata of Chordata are more closely related to vertebrates than the other invertebrate taxa studied here, we did not identify any viruses that were clearly ancestors of vertebrate-specific virus families.

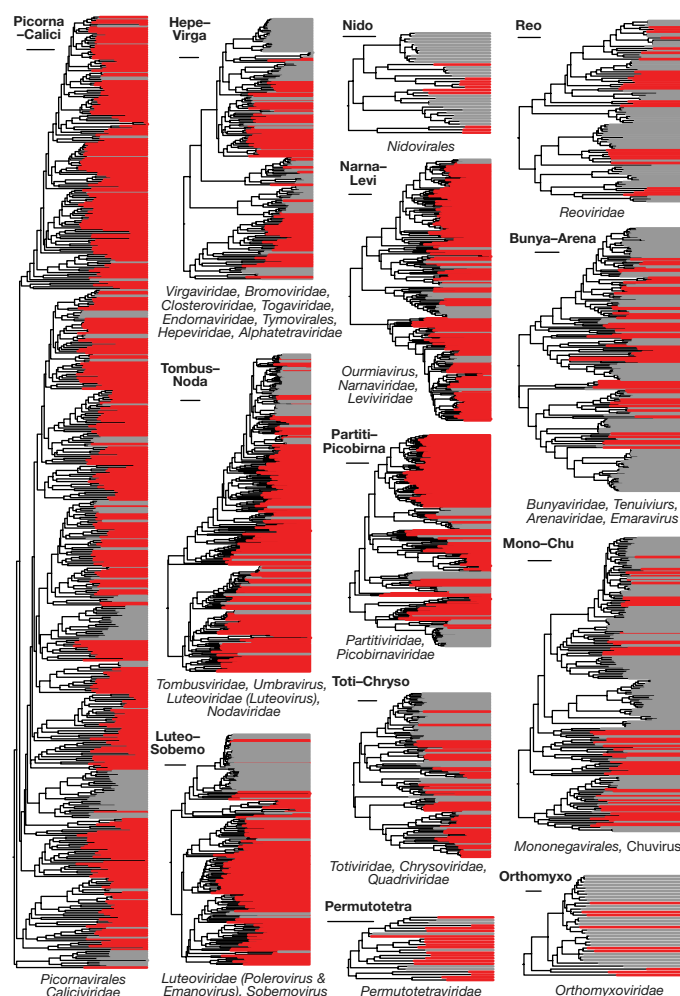


Figure 2 | Phylogenetic diversity of RNA viruses. Thirteen phylogenetic trees representing the major clades of RNA virus RdRp domains (see main text for definitions). Within each tree, the viruses discovered here are shaded red, while those described previously are shaded grey. The name of each clade is shown to the top left of each phylogeny and the names of the families or genera within clade are shown below the tree. Each scale bar indicates 0.5 amino acid substitutions per site. More detailed trees for each clade are shown in Supplementary Data 1–21, and their genome structures are shown in Supplementary Data 22–36.

Although viruses from divergent host taxa tend to form separate phylogenetic groups, suggesting that these virus–host associations have been established over long evolutionary timescales, there was generally little resemblance between the phylogenetic histories of viruses and their hosts, such that strict virus–host co-divergence cannot always be assumed. Indeed, there are clear examples of cross-species transmission of viruses among divergent host taxa. For example, several viruses that infect plants (such as *Tenuivirus* and *Fijivirus*; Supplementary Data 7 and 18, respectively) may be derived from arthropod viruses as they are nested within arthropod viruses on the phylogenies and the phylogenetic divergence between the plant and arthropod viruses was shallow.

Patterns of RNA virus genome evolution

Despite the presence of conserved RdRp sequences, the evolutionary histories of the structural and non-structural parts of the virus genomes characterized here often differed substantially (Fig. 3 and Supplementary Data 22–36). A single RdRp clade may contain coat proteins from diverse clades, and vice versa (Fig. 3 and Extended Data Table 1), indicative of widespread recombination among structural and non-structural genomic regions over long evolutionary timescales. Such incongruence is commonly observed within and between the

major groups of positive-sense RNA viruses (that is, the Tombus–Noda, Luteo–Sobemo, Hepe–Virga, Permutotetra, Astro and Narna–Levi clades). Notably, the occurrence of recombination seems to be unaffected by genome organization. There were major differences between the tree topologies of the RdRp and coat protein domains for unsegmented viruses of the ‘Tombus–Noda’ clade (Extended Data Fig. 2), and the correlation coefficient between the two genetic distance matrices for these proteins was low (Extended Data Fig. 2b). By contrast, the ‘Picorna–Calici’ clade had a more stable genome structure (Supplementary Data 33) and lower rate of genetic exchange (Extended Data Fig. 2b). Envelope glycoproteins were also involved in inter-virus recombination, although such events were rarer. Specifically, we documented recombination events involving the glycoproteins of highly divergent virus groups, including within negative-sense RNA viruses, between negative- and positive-sense RNA viruses, and even between negative-sense RNA and DNA viruses (Extended Data Table 1 and Supplementary Data 27, 28).

The data also show that the evolution of structural genes involves the gain and loss of genes, which can occur in both segmented and unsegmented viruses. We found viruses with multiple copies of structural genes, such as coat protein genes in the Hepe–Virga clade (Extended Data Fig. 3a) and glycoprotein genes in the Mono–Chu clade (Extended Data Fig. 3b). In addition, their diverse positions on the phylogeny suggest that these additional gene copies were independently acquired through lateral gene transfer rather than being generated *de novo* by gene duplication (Extended Data Fig. 3a). In other taxa, a reduction in the number of genes encoding structural proteins has been observed. Structural genes are, for example, relatively more frequently lost in negative-sense RNA viruses (Extended Data Fig. 3c), and viruses with such ‘reduced’ genomes are found in Nematoda, Arthropoda, and Platyhelminthes (Supplementary Data 27–28). Across the dataset as a whole, gene loss most often involved the glycoprotein, although several viruses within the Bunya–Arena clade may lack both glycoprotein and nucleoprotein genes (Supplementary Data 27; see Methods). Viruses with no structural proteins are also present in some positive-sense and double-stranded RNA (dsRNA) viruses, such as the *Endornaviridae*, *Hypoviridae*, *Narnaviridae* and *Umbravirus* in the Hepe–Virga, Hypo, Narna–Levi, and Tombus–Noda clades, respectively¹⁴, and we can tentatively identify clusters of viruses whose genome may only contain a replicase (Supplementary Data 23, 26, 27, 31).

We identified a number of protein domains in the non-structural part of the genome that are shared among divergent viruses, and even with cellular organisms (Extended Data Table 1). These include the RNA

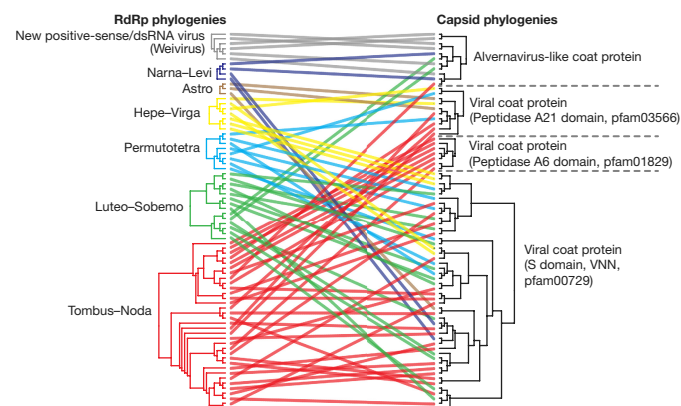


Figure 3 | Genetic exchange among RNA viruses. Comparison of the phylogenetic trees of 76 representative viral genomes with different types of structural protein (that is, four major types of capsid protein) and the equivalent phylogenies obtained for their RdRp amino acid sequences (eight clades as defined in the text, shown in different colours). Line colours correspond to those of the RdRp clade as shown to the left of the figure. Widespread recombination can be inferred when RdRp clades are associated with different types of structural protein, and vice versa.

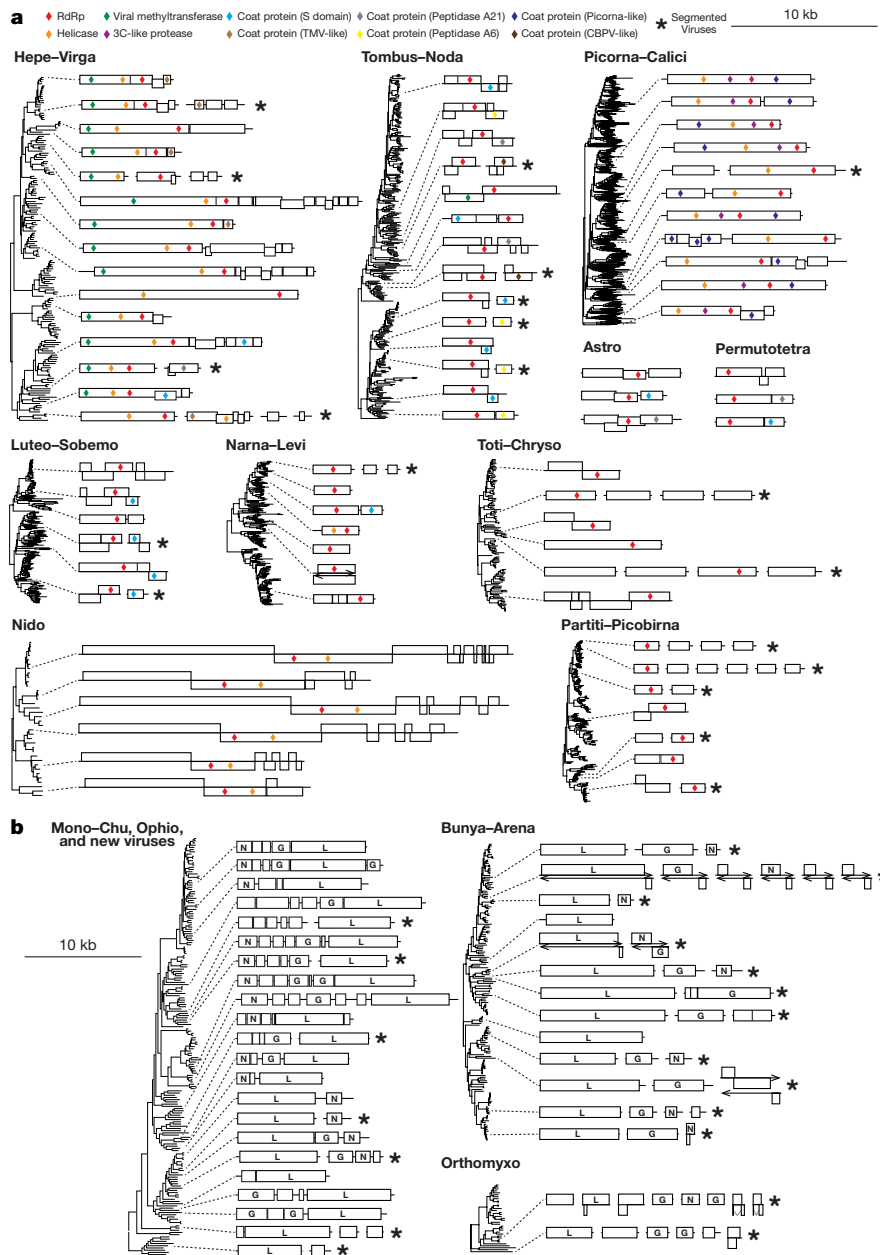


Figure 4 | Evolution of genome organization in RNA viruses. **a**, Genome evolution in ten representative clades of positive-sense and dsRNA viruses. Genome order follows the RdRp phylogeny. The genomes are drawn to a unified length scale shown at the top. The panel shows the pattern of segmentation (asterisks), the locations of major nonstructural and structural domains (coloured diamonds) and the arrangement of open reading frames. **b**, Genome evolution of negative-sense RNA viruses. N, G, and L indicate homologues from nucleoproteins, glycoproteins, and the polymerase, respectively. More detailed depictions of genome evolution are shown in Supplementary Data 22–36.

helicase, methyltransferase, exonuclease, protease, ADP-ribose binding protein (the macro domain), dsRNA binding protein, and even the *Escherichia coli* swarming motility protein (the NADAR domain). For example, we identified the parallel acquisition of eukaryotic-origin exonuclease domains (Extended Data Fig. 4a). Notably, the virus genomes that contain these exonuclease domains were highly divergent, although all were present in the same host (*Ligia exotica*, Extended Data Fig. 4a). In the case of the serine protease, virus diversity appeared both within and outside of the phylogenetic diversity of cellular proteins (Extended Data Fig. 4b), indicative of independent gene acquisitions. Also of note was the wide, but highly sporadic, phylogenetic distributions of some of these domains (such as Macro and NADAR; Extended Data Table 1) and the variable insertion locations in the genome (Extended Data Fig. 4c), again pointing to multiple, independent gene acquisitions.

There have been major reconfigurations of viral genome organization through evolutionary history, including the number and arrangement of open reading frames, the order of structural and non-structural genes, and the occurrence and extent of segmentation (Fig. 4 and Supplementary Data 22–36). These features, which are often regarded as conservative traits, in reality show great flexibility at the deep

evolutionary scale studied here. Examination of the newly discovered viruses reveals that the evolution of segmented genomes, or the loss of segmentation, has occurred frequently. Several viruses experience high frequencies of segmentation, including members of the Tombus–Noda and Mono–Chu clades, in which the break-up and reunification of structural and non-structural genes has occurred relatively frequently (Fig. 4 and Supplementary Data 28, 34). The dynamics of segmentation were also reflected by changes in the number of segments. For instance, the number of segments in both the Partiti–Picobirna and Bunya–Arena clades vary from 1 to 6 (Fig. 4). Notably, the change in segment numbers is not only associated with gene break-up and unification, but also with the gain and loss of genes. We also found a tri-segmented counterpart of the normally bi-segmented arenaviruses that exhibited no structural gene homology with other members of this group (Supplementary Data 7, 27). Despite this flexibility, it is also the case that a genomic plan comprising multiple segments can be conserved for extended time periods, as is seen in the reoviruses and orthomyxoviruses. In one instance in the latter, we identified a highly divergent virus from the earthworm that possessed a similar genomic plan to other orthomyxoviruses (with 6 segments), including those from vertebrates (Fig. 4 and Supplementary Data 29).

Discussion

We have used a simple, yet powerful, metagenomic approach to characterize the viromes of diverse invertebrates. This approach is relatively unbiased, as no attempt is made to enrich viral particles through filtering, centrifugation and nuclease treatment. Although some invertebrates seemingly harboured a high proportion of viral RNA, which evidently depends on the infection status in the host in question, we were unable to determine whether the viruses identified here have any impact on host biology, including as agents of disease. Despite this, it is clear that for many invertebrates infection by multiple RNA viruses is likely to be the norm rather than the exception^{15–17}.

The whole-transcriptome approach employed here allowed us to characterize the virome of a diverse array of invertebrates, providing a new perspective on viral biodiversity. A number of virus families that were previously only known to infect plants, fungi, and protists are now visible in invertebrates. Viruses infecting divergent phyla were dispersed throughout the phylogenetic trees and exhibited diverse patterns of clustering, reflecting a complex interplay between long-term virus–host associations, including co-divergence, as well as frequent host jumping¹⁸.

Despite the relatively high frequency of potential cross-species virus transmission documented here, there were also probable cases of long-term virus–host co-divergence. The viruses found in several species of parasitic nematodes tend to form monophyletic clusters, within which the phylogeny of viruses mirrors that of their hosts. In addition, the genetic diversity of RNA viruses within the Narnia–Levi clade can be placed into three groups: those that infect bacteria (leviviruses), those that infect mitochondria (that is, mitoviruses that utilize the mitochondrial genetic code), and those that infect other organisms. This separation could in theory have occurred when the α -proteobacteria became intracellular symbionts^{19,20}. Finally, despite the massive expansion of virus diversity documented here, some vertebrate viruses (such as those from the *Picornaviridae*, *Paramyxoviridae* and *Hepeviridae* families) remain monophyletic, with viruses from mammals, birds, reptiles, and fish occupying similar phylogenetic relationships to those of their hosts groups, probably indicative of long-term co-divergence.

We have necessarily inferred phylogenetic trees using the relatively conserved gene encoding RdRp. However, the evolutionary history of the entire genome is evidently more complex and not necessarily consistent with that of the RdRp domain. Indeed, at deep evolutionary timescales it is easier to trace the evolutionary history of individual functional units, such as the RdRp, helicase, and capsid, rather than that of intact viral genomes. Such modular genome evolution is reflected in three aspects of genetic diversity. First, there is great flexibility in the organization of functional units, including changes in genome segmentation and gene order. Second, different functional ‘units’ within genomes can be acquired or removed independently, although such processes occur relatively infrequently, which is likely to reflect strong restrictions on virus genome size. Thus, the simplest virus genome may contain only the replication module, whereas the most complex can contain multiple structural units or accessory units. Third, there is clear evidence for the exchange of functional units among viruses, particularly for structural proteins that can seemingly move large phylogenetic distances. Hence, the macroevolution of RNA viruses parallels the modular evolution previously proposed for bacteriophages²¹, albeit with differences in timescale and mechanism. However, in the face of such abundant diversity, it is notable that none of the RNA viruses described here has a genome that exceeds the previously defined maximum of approximately 32 kb, probably reflecting intrinsic size constraints owing to error-prone replication¹⁰.

By sampling a diverse range of invertebrate taxa, we have revealed unprecedented levels of RNA virus genetic diversity that both re-shapes our understanding of the patterns and processes of their evolution and highlights the limitations of our knowledge on what are likely to be the most abundant organisms on earth²². A full understanding of virus evolution and ecology will require an extensive survey of diverse host organisms using the types of metagenomic approach outlined here.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 May; accepted 17 October 2016.

Published online 23 November 2016.

- Koonin, E. V., Senkevich, T. G. & Dolja, V. V. The ancient Virus World and evolution of cells. *Biol. Direct* **1**, 29 (2006).
- Junglen, S. & Drosten, C. Virus discovery and recent insights into virus diversity in arthropods. *Curr. Opin. Microbiol.* **16**, 507–513 (2013).
- Li, C. X. *et al.* Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife* **4**, e05378 (2015).
- Bekal, S., Domier, L. L., Niblack, T. L. & Lambert, K. N. Discovery and initial analysis of novel viral genomes in the soybean cyst nematode. *J. Gen. Virol.* **92**, 1870–1879 (2011).
- Ballinger, M. J., Bruenn, J. A., Hay, J., Czechowski, D. & Taylor, D. J. Discovery and evolution of bunyaviruses in arctic phantom midges and ancient bunyavirid-like sequences in insect genomes. *J. Virol.* **88**, 8783–8794 (2014).
- Qin, X. C. *et al.* A tick-borne segmented RNA virus contains genome segments derived from unsegmented viral ancestors. *Proc. Natl Acad. Sci. USA* **111**, 6744–6749 (2014).
- Tokarz, R. *et al.* Virome analysis of *Amblyomma americanum*, *Dermacentor variabilis*, and *Ixodes scapularis* ticks reveals novel highly divergent vertebrate and invertebrate viruses. *J. Virol.* **88**, 11480–11492 (2014).
- Webster, C. L. *et al.* The discovery, distribution, and evolution of viruses associated with *Drosophila melanogaster*. *PLoS Biol.* **13**, e1002210 (2015).
- Shi, M. *et al.* Divergent viruses discovered in arthropods and vertebrates reveal the evolutionary history of the *Flaviviridae* and related viruses. *J. Virol.* **90**, 659–669 (2015).
- Holmes, E. C. *The Evolution and Emergence of RNA Viruses*. (Oxford Univ. Press, 2009).
- Koonin, E. V. The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J. Gen. Virol.* **72**, 2197–2206 (1991).
- Feschotte, C. & Gilbert, C. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat. Rev. Genet.* **13**, 283–296 (2012).
- Philippe, H., Lartillot, N. & Brinkmann, H. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol. Biol. Evol.* **22**, 1246–1253 (2005).
- King, A. M. Q., Adams, M. J., Carstens, E. B. & Lefkowitz, E. J. *Virus Taxonomy: 9th Report of the International Committee on Taxonomy of Viruses*. (Elsevier Academic Press, 2012).
- Gauthier, L. *et al.* Viral load estimation in asymptomatic honey bee colonies using the quantitative RT-PCR technique. *Apidologie (Celle)* **38**, 426–435 (2007).
- Genersch, E. *et al.* The German bee monitoring project: a long term study to understand periodically high winter losses of honey bee colonies. *Apidologie (Celle)* **41**, 332–352 (2010).
- Tentcheva, D. *et al.* Prevalence and seasonal variations of six bee viruses in *Apis mellifera* L. and *Varroa destructor* mite populations in France. *Appl. Environ. Microbiol.* **70**, 7185–7191 (2004).
- Baranowski, E., Ruiz-Jarabo, C. M. & Domingo, E. Evolution of cell recognition by viruses. *Science* **292**, 1102–1105 (2001).
- Andersson, S. G. & Kurland, C. G. Origins of mitochondria and hydrogenosomes. *Curr. Opin. Microbiol.* **2**, 535–541 (1999).
- Gray, M. W., Burger, G. & Lang, B. F. Mitochondrial evolution. *Science* **283**, 1476–1481 (1999).
- Botstein, D. A theory of modular evolution for bacteriophages. *Ann. NY Acad. Sci.* **354**, 484–490 (1980).
- Suttle, C. A. Viruses in the sea. *Nature* **437**, 356–361 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements This study was supported by the National Natural Science Foundation of China (Grants 81290343, 81273014, 81672057), the Special National Project on Research and Development of Key Biosafety Technologies (Grants 2016YFC1201900, 2016YFC1200101), the 12th Five-Year Major National Science and Technology Projects of China (2014ZX10004001-005), and an NHMRC Australia Fellowship (GNT1037231).

Author Contributions Conceptualization: M.S. and Y.-Z.Z. Methodology: M.S., L.-J.C., C.-X.L., J.L., J.-S.E., J.B., E.C.H. and Y.-Z.Z. Investigation: M.S., X.-D.L., J.-H.T., L.-J.C., X.C., C.-X.L. and X.-C.Q. Writing (original draft): M.S., E.C.H. and Y.-Z.Z. Writing (review and editing): M.S., X.-D.L., J.-H.T., L.-J.C., X.C., C.-X.L., J.-S.E., J.X., E.C.H. and Y.-Z.Z. Funding Acquisition: J.X., E.C.H. and Y.-Z.Z. Resources (sampling): M.S., X.-D.L., J.-H.T., L.-J.C., X.C., C.-X.L., J.-P.C., W.W. and Y.-Z.Z. Resources (computational): M.S., J.L., J.B. and E.C.H. Supervision: E.C.H. and Y.-Z.Z.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.-Z.Z. (zhangyongzhen@icdc.cn).

Reviewer Information Nature thanks E. Ghedin, D. Obbard and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. These experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Sample collection and processing. This study was based on the analysis of 87 libraries of invertebrate samples obtained from various locations in China (Supplementary Table 1). The sampling included land and freshwater organisms from Anhui, Beijing, Hubei, Xinjiang and Zhejiang provinces. Marine and coastal samples were obtained from Zhejiang (East China Sea) and Guangxi (South China Sea) provinces.

A total of 54 of the libraries were from arthropods (phylum Arthropoda). Of these arthropod libraries, 19 have been described previously^{3,6,9}, with 35 additional libraries newly obtained here. We sampled across all four subphyla from the phylum Arthropoda: (i) for the subphylum Chelicerata we sampled from the class Merostomata (horseshoe crabs) and Arachnida (spiders and ticks); (ii) for the subphylum Crustacea we sampled from the classes Branchiopoda (water fleas), Maxillopoda (barnacles), and Malacostraca (crabs, shrimps, crayfish, woodlice, wharf roaches, and so on); (iii) for the subphylum Hexapoda we sampled from nine orders (Coleoptera, Dermaptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera, Odonata, Orthoptera, and Siphonaptera), all within the class Insecta; and (iv) for the subphylum Myriapoda we sampled from the classes Chilopoda (centipedes) and Diplopoda (millipedes).

All of the taxa from the phyla Nematoda and Platyhelminthes sampled here were parasites. The majority of the nematodes were from the family Ascaridida (class Secernentea) and found in the stomachs of pigs (libraries HC and WHZHC) and birds (libraries HC, JMYJCC, SYJFC, and WHJHC). We also collected three divergent nematode species within the class Secernentea from the stomachs of mice (library WHSWHC), the thoracic cavity of birds (library XSNXC), and the stomachs of snakes (library XZSJSC). In addition, we included a nematode species that infects mosquitoes (library WCLXSC) tentatively classified in the genus *Romanomermis* (class Adenophorea). For the phylum Platyhelminthes, our sampling comprised one species of tapeworm (*Taenia* sp.) discovered in the liver of rodents, as well as two species of blood flukes (*Schistosoma mansoni* and *S. japonicum*) initially obtained from an intermediate host (*Oncomelania*) and then matured in rabbits.

We also sampled three phyla within the superphylum Lophotrochozoa, namely Mollusca, Sipuncula, and Annelida. Our samples from the phylum Mollusca comprised a number of marine and freshwater species, including the classes Bivalvia (such as clams, mussels, oysters), Gastropoda (Chinese land snails, various sea snails, etc.), and Cephalopoda (octopus). For the phylum Sipuncula (peanut worms), we sampled the most common species in the shallow waters of the South China Sea, namely *Phascolosoma esculenta* and *Sipunculus nudus*. For the phylum Annelida, we sampled three representative classes: Polychaeta (sandworms), Oligochaeta (earthworms), and Hirudinea (leeches).

All the samples described above were protostomes. For the deuterostomes our samples comprised the phyla Chordata and Echinodermata. Within the Chordata, we sampled two species from the subphylum Tunicata, the invertebrate group most closely related to the vertebrates. Within the Echinodermata, we sampled representatives from the classes Echinoidea (sea urchins) and Holothuroidea (sea cucumbers). Finally, we examined one species from the radially symmetric phylum Cnidaria (sea anemones) as a representative of basal lineages within the Metazoa.

All samples were captured alive and stored at -80°C . A proportion of the animals were left in Petri dishes (including pillworms, earthworms, centipedes, fiddler crabs, hermit crabs, woodlice, sandworms, oysters, Paphia shells, razor shells, Murex snails, Turritella sea snails, and Chinese land snails) or in purified sea water (horseshoe crabs, penaeid shrimps and mantis shrimp) for up to 24 h to clean stomach contents before transferring to -80°C . Sample processing often involved the entire animal. However, for those animals with large body sizes, hard shells, or tissues that were difficult to homogenize, dissection was performed to obtain the entire inner organs (visceral mass for molluscs) or parts of different inner organs (Supplementary Table 1). During the dissection, the content of gut was intentionally excluded to reduce contamination.

Host species identification was initially carried out by experienced field biologists. Further confirmation was based on analysing the cytochrome *c* oxidase subunit I (COI) gene. COI sequences were first obtained from the assembled contigs and then by Sanger sequencing. They were subsequently compared against the NCBI non-redundant nucleotide database and the BOLD database (<http://www.boldsystems.org/>) for host species identification and confirmation.

RNA library construction and sequencing. On the basis of the complexity of the component samples, our libraries were divided into two categories: (i) simple libraries that contained single or multiple individuals from one or two closely

related species; (ii) mixed libraries that contained multiple species from a particular taxonomic group. For example, the library WLJQ is a mix of individuals from the order Decapoda sampled in the East China Sea. For some of the mixed libraries, we later sequenced individual species (that is, before pooling) to assist genome characterization.

To construct each library, the processed samples were first washed with a standard, sterile, RNA and DNA-free PBS solution (GIBCO). This washing was performed three times, and each time the solution was pipetted to agitate the solution and remove the surface organisms/material while keeping the organisms/tissue intact. The samples were then homogenized in 500–700 μl PBS solution using the Mixer mill MM400 (Retsch). Total RNA was extracted using TRIzol LS reagent (Invitrogen) and subsequently purified using EZNA Total RNA Kit (OMEGA). Aliquots of the resultant RNA solutions were then pooled in equal quantity and quality checked using an Agilent 2100 Bioanalyzer (Agilent Technologies) before library construction and sequencing. For most libraries we used the TruSeq total RNA Library Preparation protocol (Illumina). rRNA was removed using either the Ribo-Zero-Gold (Human–Mouse–Rat) Kit (Illumina) or the Ribo-Zero-Gold (Epidemiology) Kit (Illumina). For five libraries we used the TruSeq mRNA Library Preparation protocol (Illumina) that only targeted RNA with poly(A) tails, although these sequencing results were not used in the quantification. The information on library construction methods for each pool can be found in Supplementary Table 1. Paired-end (90 or 100 bp) sequencing of each RNA library was performed on the HiSeq 2000 platform (Illumina). All library preparation and sequencing was carried out by BGI Tech.

Sequence assembly and RNA virus discovery. For each library, sequencing reads were quality trimmed and assembled *de novo* using the Trinity program²³ with default parameter settings. No filtering of host/bacterial reads was performed before the assembly. The assembled contigs were first compared (using blastx) against the database of all reference RNA virus proteins downloaded from GenBank, which include those within the taxonomic classes ssRNA viruses (txid 439488), dsRNA viruses (txid 35325), and *Deltavirus* (txid 39759). We set the *e*-value to 1×10^{-5} to maintain high sensitivity and a low false-positive rate. To detect highly divergent viruses, we performed domain-based blast by comparing the assembled contigs against the Conserved Domain Database (CDD) version 3.14 with an expected value threshold of 1×10^{-2} . Sequences with positive hits to the domain RNA_dep_RNAP (cd01699) were retained. After the initial screening, potential false-positives were discovered by (i) comparing (blastx) putative viral contigs against the entire non-redundant protein database, and (ii) inspecting the sequence alignment for conserved domains. The quality-filtered viral sequences were incorporated into the reference protein database for a second round of blastx.

To identify potential retroviruses and retrotransposons, we examined the domain blast results for any hit to the superfamily reverse-transcriptase-like domain (RT_like, cl02808), excluding those related to RNA_dep_RNAP (cd01699). To avoid false positives, we used a higher *e*-value threshold (1×10^{-5}). All putative reverse transcriptases recovered were aligned to related proteins to determine the presence of key motifs.

Confirmation and extension of virus genomes. Viral contigs with unassembled overlaps or from the same scaffold were merged using the SeqMan program implemented in the Lasergene software package v7.1 (DNASTar). Gaps were filled by RT-PCR and Sanger sequencing. To confirm the assembly results, reads were mapped back to the full length genome with Bowtie2 (ref. 24) and inspected using the Integrated Genomics Viewer²⁵. For genomes with novel structures or that contained sequences originating from lateral gene transfer events, we verified the complete or near complete viral genome by designing overlapping primers based on the assembled sequences (Supplementary Table 2). To check these viruses have no DNA stage, we used PCR and Sanger sequencing to examine the DNA extracted from the same set of viruses on which we performed RNA genome confirmation (Supplementary Table 2). Finally, genome termini were determined by RNA circularization or 5'/3' RACE kits (TaKaRa) as described previously³.

Transcriptome annotation. For each library, we annotated the top 1,000 most common transcripts. The quantity of the transcripts were determined using the RSEM program²⁶ implemented in Trinity. The top 1000 common transcripts were then compared against four databases: (i) the non-redundant protein database (nr), (ii) the non-redundant nucleotide database (nt), (iii) the whole-genome shotgun database (wgs), and (iv) the Conserved Domain Database (CDD). The resultant information was used to identify the origin of sequences. Host and mitochondrial genes were identified using two criteria: (a) well-characterized domain/functional information from the domain-based blast results or, in the case of non-coding sequences, from the results of the blastn search against the nt database, and (b) the presence of identical sequences in the host genome or homologous genes in the genome of related host taxa. RNA virus genomes and retrotransposons were identified as described under the 'Sequence assembly and RNA virus discovery'

section. Bacterial contigs were identified if they exhibited high nucleotide similarity (>80%) to a particular bacterial genome. Finally, the transcripts of DNA viruses were identified if they shared protein sequence similarity with either viral polymerases or virus-specific genes (for example, capsid proteins) described previously. The remaining contigs were tentatively annotated as 'undetermined'.

Determination of additional virus genome segments. In those viruses with multiple segments we used various strategies to search for genome segments other than the RdRp. The majority of such segments were found based on their homology to the proteins of related reference viruses. For segments that encode proteins with no known homologues, we used *in silico* approaches that collectively utilize information on RNA quantity, protein structure, and/or conserved genome termini. To determine that these segments belonged to the same virus, we checked: (i) the sequencing depth of the segments; (ii) the presence of inverted complementary genome termini or conserved regulatory sequences in non-coding regions of the genome; (iii) whether the segments were found in the same samples; and (iv) the phylogenetic positions of related viral proteins. For example, Wuhan cricket virus 2 had six potential segments, of which only two exhibited homology to known viruses. The remaining segments were initially identified as 'undetermined' protein-coding contigs which, like the segments encoding the RdRp and capsid, were the most frequent contigs in the transcriptome. Further alignment of the six segments revealed conserved stretches at both the 5' and 3' ends (confirmed by RACE), implying they are derived from the same virus. Similarly, the five unknown segments of Changping earthworm virus 2, a divergent member of the 'Orthomyxo' clade, were identified by RNA quantity and the presence of the same inverted complementary genome termini. In addition, two of the segments were identified as potential glycoprotein genes because both encoded proteins had an N-terminal signal domain, a C-terminal or mid-point transmembrane domain and putative glycosylation sites.

Despite our best efforts to address the lack of sequence similarity, it remained difficult to fully characterize the genome segments in a number of divergent viruses. In some of these cases the viruses were likely to be unsegmented even though related viruses appeared to harbour multiple segments. For example, in the case of the nematode-associated lineage within the 'Bunya-Arena' clade (that is, that including Shayang *Ascaridia galli* virus 1), our annotation of contigs with similar abundance levels suggested they were either of host origin or derived from other viruses, rather than representing another protein-coding segment.

It is, however, critical to acknowledge that all segment identification is tentative at this stage, and requires confirmation from virus isolation.

Estimation of viral transcript frequency. To help determine the frequency of viral RNAs, we estimated the percentage of reads that mapped to viral RNA within the transcriptome of each host. To reduce any bias caused by the unequal efficiency of rRNA removal during library preparation, we first removed reads that mapped to rRNA contigs from each library. The remaining reads were then mapped to the entire collection of virus sequences within the library, from which we calculated the overall percentage of viral reads. The proportion of individual viral RNAs was then estimated based on the mapping results. To confirm those results in which viral RNA comprised a large percentage of transcripts within the host transcriptome, we re-extracted the total RNA from aliquots of the original homogenates and performed RNA-seq library preparation and sequencing as described above.

Inference of virus evolutionary history and virus nomenclature. To infer the phylogenetic relationships among RNA viruses we collected all replicase proteins translated from the virus sequence collections described above. For comparison, we downloaded from GenBank reference virus genomes from all established families and floating genera of RNA viruses (excluding retro-transcribing viruses) and non-reference virus genomes not included in the current classification scheme but that are relatively closely related to the viruses discovered here. The viral replicase sequences were then aligned using MAFFT version 7 employing the E-INS-i algorithm²⁷. All alignments were trimmed so that they only contained the RdRp and its neighbouring conserved domains. All ambiguously aligned regions were then removed using the TrimAl program²⁸. For each sequence alignment, the best-fit model of amino acid substitution was determined using ProtTest 3.4 (ref. 29). Phylogenetic trees were then inferred using the maximum likelihood approach (ML) implemented in PhyML version 3.0 (ref. 30), employing Subtree Pruning and Regrafting (SPR) branch-swapping. Branch support was accessed using an approximate likelihood ratio test (aLRT) with the Shimodaira-Hasegawa-like procedure as implemented in PhyML.

The (provisional) naming of viruses was based on the following approach: the name of a virus characterized by a high proportion of RNA transcripts (>0.1% of non-rRNA reads) contains information on the geographic location of sampling,

the host common name (such as 'tick'), and a virus number; whereas the name of a virus characterized by a low proportion of RNA transcripts (<0.1% non-rRNA reads), for which host assignments are less certain, contains information on geographic region of origin, closest family/genus (for example, 'astro-like'), and virus number. The strain name of each virus (shown in each detailed tree, Supplementary Data 1–21) comprises the library abbreviation followed by its contig number.

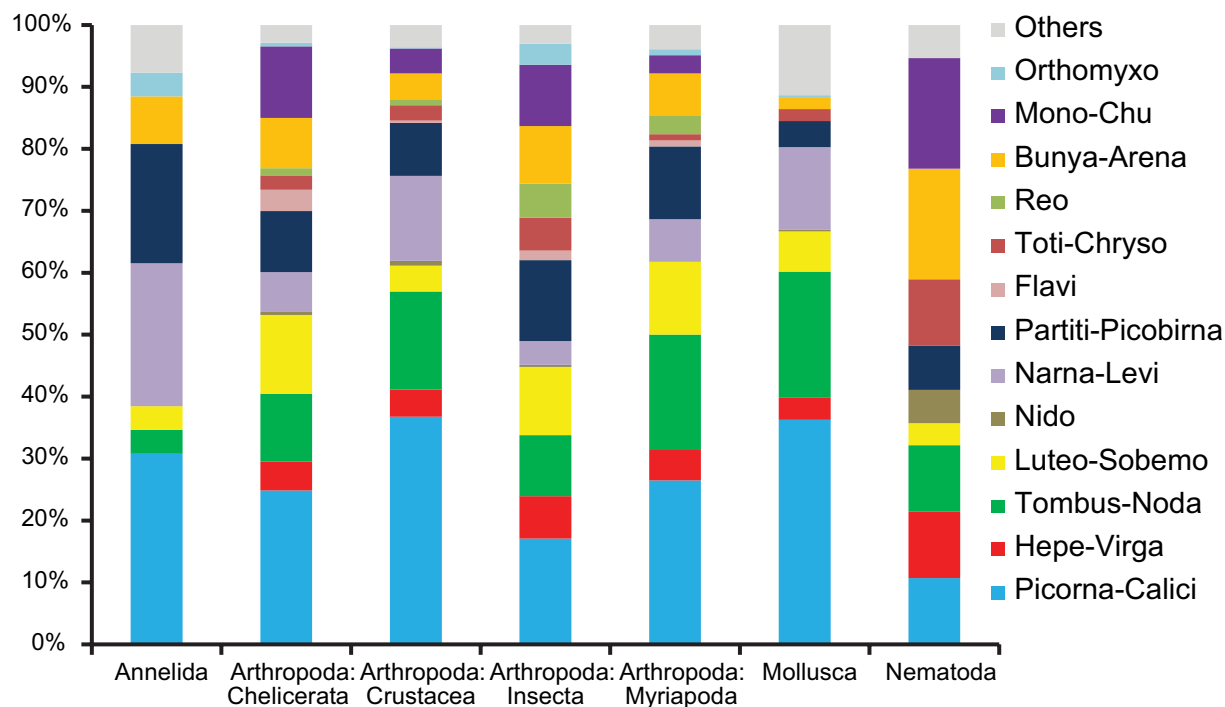
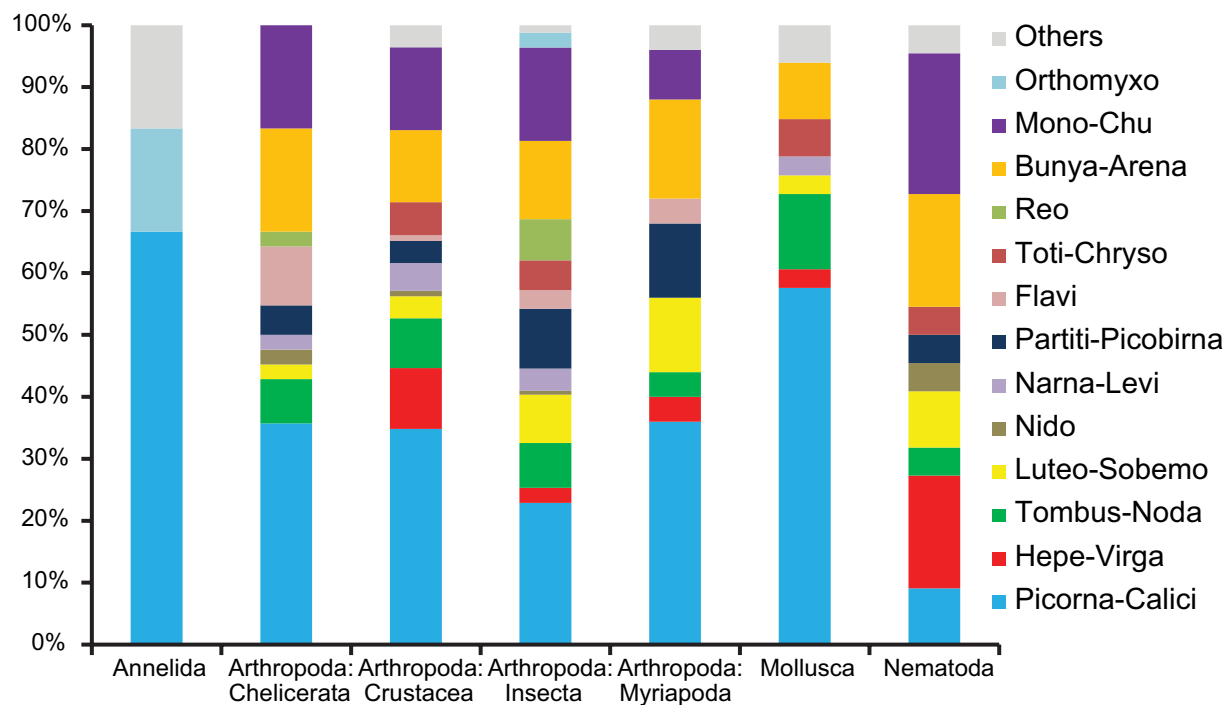
Virus genome annotation. Potential viral open reading frames (ORFs) were predicted based on two criteria: (i) the predicted amino acid sequences were longer than 100 amino acids in length, and (ii) if a short ORF (<200 amino acids) was completely nested within a larger one it was not regarded as a potential ORF unless it had a homologue in a closely related virus. The annotations of these ORFs were mainly based on comparisons to the Conserved Domain Database. Domains that were potentially subject to lateral gene transfer were further examined by sequence alignment and phylogenetic analyses. For the remaining ORFs, we predicated their potential functions by blast searches against the nr protein database with an *e*-value threshold of 1×10^{-5} and by primary protein structure predication using the programs SignalP, TMHMM, and NetNGlyc available through the website (<http://www.cbs.dtu.dk/services/>).

Analysis of recombination and lateral gene transfer. For each newly identified protein, we searched for any potential homologues against all RNA virus proteins (including those newly identified here), all DNA virus proteins, and those from the cellular organisms (using a subset of the nr database). We also identified homology if the proteins matched the same domain in the structure-based blast. On the basis of these results, we identified several well-established homologous protein clusters. We then mapped these protein clusters onto the RdRp phylogenies, which enabled us to identify topological inconsistencies that were likely to be the result of lateral gene transfer. To identify homologous recombination events, we compared the phylogenies of each homologous protein cluster to that of the RdRp. To measure the degree of phylogenetic incongruence, we transformed the two phylogenies into patristic genetic distance matrices and calculated the Pearson correlation coefficient.

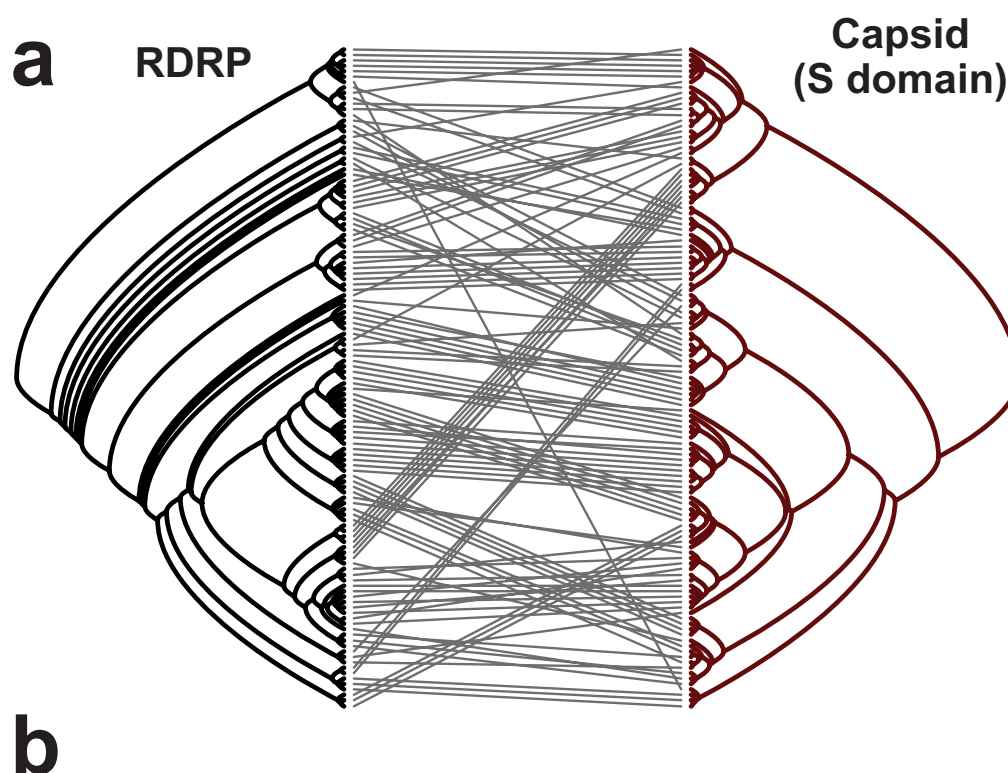
Analysis of endogenous virus elements. All genomes from cellular organisms available in GenBank were downloaded and incorporated into our local genome nucleotide database. Endogenous copies of the exogenous viruses (that is, EVEs) were detected using the tblastn algorithm against this database. The query involved amino acid sequences translated from both the virus genomes newly identified here as well as the reference virus genomes used in this study. The threshold for the search was set to 100 amino acids for length and 1×10^{-20} for *e*-value. For each potential endogenous virus, the query process was reversed to determine their corresponding phylogenetic group. The results were also checked manually to exclude those sequences involved in lateral gene transfer.

Data availability. All new sequence reads generated here are available at the NCBI Sequence Read Archive (SRA) database under the BioProject accession PRJNA318834 (Supplementary Table 1). All virus genome sequences generated in this study have been deposited in GenBank under the accession numbers KX882764–KX884872 (Supplementary Table 2). All viruses discovered in this study (fasta format), sequence alignments (fasta format), and phylogenetic trees (newick format) are available at https://figshare.com/articles/Redefining_the_invertebrate_RNA_virosphere/3792972.

23. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
24. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
25. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
26. Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A. & Dewey, C. N. RNA-seq gene expression estimation with read mapping uncertainty. *Bioinformatics* **26**, 493–500 (2010).
27. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
28. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
29. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165 (2011).
30. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).

a**b**

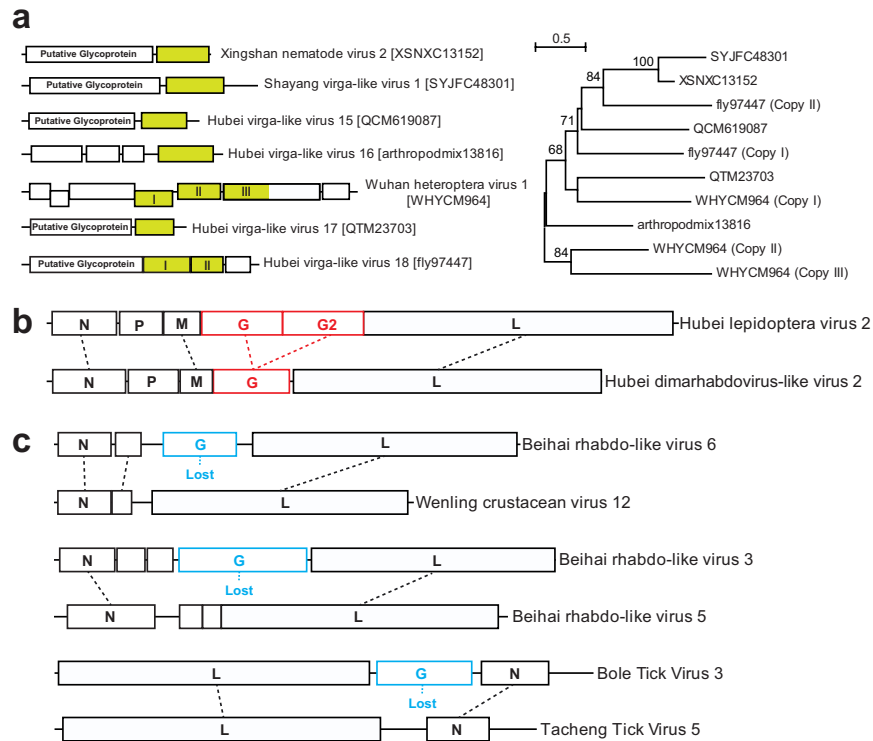
Extended Data Figure 1 | The contribution of major viral clades to the total virome of each host phylum/order. a, b, These analyses are based on viruses at all frequency levels (a), and viruses in which the frequency exceeds 0.1% of the total number of non-rRNA reads (b).



RdRp	Structural Protein	Correlation
Tombus-noda viruses (only unsegmented)	S domain	0.286
Aquatic Picorna-like Cluster	Picorna-like capsid	0.712
<i>Dicistroviridae</i> -related viruses	Picorna-like capsid	0.695
<i>Dimarhabdovirus</i> group	Nucleoprotein	0.725
Chuviruses and relatives (only unsegmented)	Chuvirus-like glycoprotein	0.559
New -ve RNA virus (Qinvirus)	Putative nucleoprotein	0.231
New +ve/ds RNA virus (Weivirus)	<i>Alvenoviridae</i> -like capsid	0.010

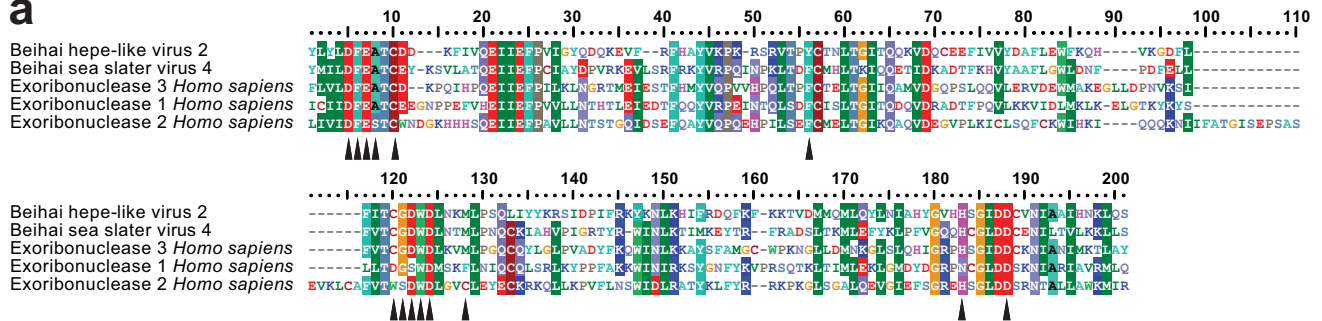
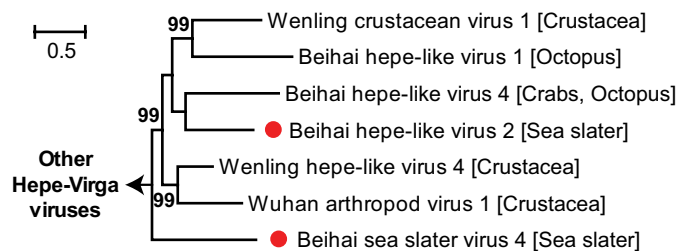
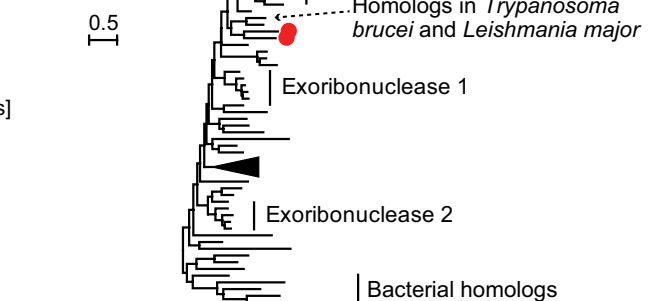
Extended Data Figure 2 | Phylogenetic incongruence between the RdRp and structural proteins. **a**, Match between the phylogenies of the RdRp and coat proteins (S-domain like) for non-segmented members of the Tombus–Noda clade. The relationship between the two phylogenies

is displayed to maximize topological congruence. **b**, The degree of phylogenetic incongruence for different pairs of structural and non-structural phylogenies. The comparisons were based on patristic distances matrices derived from the phylogenies.

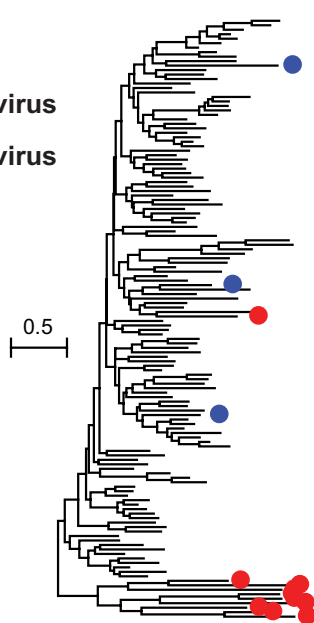


Extended Data Figure 3 | The gain and loss of RNA virus structural proteins. **a**, The parallel acquisition of multiple copies of structural proteins by viruses within the Hepe–Virga clade. Left panel shows an outline of the structural part of their genomes, with homologous structural genes marked in yellow and multiple copies of these proteins within the same genome labelled as ‘I’, ‘II’, and ‘III’. Right panel shows a maximum-likelihood phylogeny depicting the evolutionary history of the corresponding structural proteins of these viruses. **b**, Acquisition

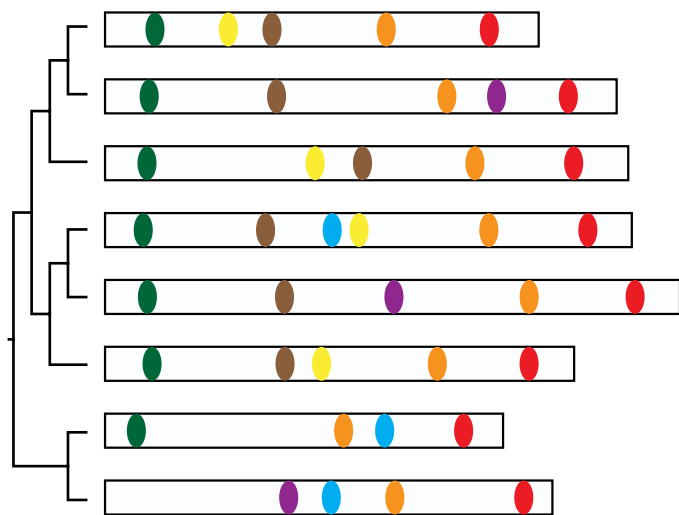
of a glycoprotein in the genome of Hubei Lepidoptera virus 2 from the Mono–Chu Clade. Its genome is compared against that of a closely related virus (Hubei dimarhabdovirus-like virus 2). Homologous proteins are connected with dotted lines, and the target glycoprotein is shown in red. **c**, Three examples of glycoprotein loss in the Mono–Chu Clade. Homologous proteins are connected with dotted lines, and the target glycoproteins are shown in blue.

a**RdRp****Exonuclease****b**

● DNA virus
● RNA virus

**c**

● RdRp ● FstJ ● Macro ● NADAR
● Helicase ● OTU ● Viral methyltransferase



Extended Data Figure 4 | Lateral gene transfer between RNA viruses and cellular organisms. **a**, Evolutionary origin of two exoribonucleases (cd06133) in two sea-slater-associated viruses (Beihai hepe-like virus 2 and Beihai sea slater virus 4). Top, alignment of viral and (human) cellular exoribonucleases. The solid triangles indicate the key catalytic sites. Lower left panel shows the phylogenetic positions of the two viruses (marked with solid red circles) whose genomes contain these exoribonucleases. The host information for each virus is shown in parentheses. Lower right panel shows the phylogenetic position of the virus exoribonucleases (solid red circle) in the context of cellular exoribonucleases. **b**, Evolutionary origin of viral serine proteases (cd00190). The phylogeny contains serine

proteases from RNA viruses (solid red circles), DNA viruses (solid blue circles) and cellular organisms. Serine proteases from RNA viruses are either highly divergent or group within the diversity of cellular proteins. **c**, Relative positions of different protein domains in the replicase of selected Hepe-Virga viruses. The domains are shown as ovals and marked with different colours, and comprise: RdRp (cd01699), Helicase (pfam01443), FstJ (pfam01728), OTU (OTU-like cysteine protease, pfam02338), Macro (cl00019), NADAR (cd15457), and viral methyltransferase (pfam01660). More detailed depictions of lateral gene transfer can be found in Supplementary Data 22–36.

Extended Data Table 1 | Distribution of homologous protein clusters across divergent taxonomic groups (RNA viruses, DNA viruses and cellular organisms)

Category	Protein (Domain)	CDD accession	Astro	Flavi	Hepe-Virga	Luteo-Sobemo	Nama-Levi	Nido	Permutotetra	Picorna-Calici	Reo	Tombus-Noda	Toti-Chryso	New +ve RNA Viruses	-ve RNA Viruses	DNA Virus	Cellular Organisms
Structural	Capsid (S domain)	pfam00729	X		X	X	X		X			X		X		X	
	Capsid (peptidase A21)	pfam03566	X		X				X			X					
	Capsid (Alvernavirus core like)	N/A				X	X					X		X			
	Glycoprotein (Okavirus like)	N/A						X							X		
	Glycoprotein (Ferak virus like)	N/A			X										X		
	Glycoprotein (Hemagglutinin-neuraminidase)	pfam00423						X							X		
	Glycoprotein (<i>Choristoneura rosaceana</i> alphabaculovirus GP64)	N/A													X	X	
	Glycoprotein (Ostreid herpesvirus 1 ORF68)	N/A													X	X	
Non-structural	RNA Helicase (picorna-like)	pfam00910					X			X							
	Viral methyltransferase	pfam01660			X							X					
	FtsJ-like methyltransferase	pfam01728		X	X										X		X
	Macro (ADP-ribose binding)	cl00019		X	X			X		X					X		X
	Ribonuclease III (dsRNA binding)	cl00054							X	X	X	X	X				X
	3'-5' exonucleases	cd06133			X												X
	E. coli swarming motility protein (NADAR)	cd15457			X			X		X							X
	OTU-like cysteine protease	pfam02338		X	X										X		X
	2OG-Fe(II) oxygenase	pfam13532			X					X		X				X	X
	Trypsin-like serine protease	cl21584	X	X	X	X		X		X						X	X
	RNA 2'-phosphotransferase	pfam01885								X							X

Persistent microbiome alterations modulate the rate of post-dieting weight regain

Christoph A. Thaiss^{1*}, Shlomik Itav^{1*}, Daphna Rothschild^{2,3*}, Mariska T. Meijer¹, Maayan Levy¹, Claudia Moresi¹, Lenka Dohnalová¹, Sofia Braverman¹, Shachar Rozin¹, Sergey Malitsky⁴, Mally Dori-Bachash¹, Yael Kuperman⁵, Inbal Biton⁵, Arie Gertler⁶, Alon Harmelin⁵, Hagit Shapiro¹, Zamir Halpern^{7,8}, Asaph Aharoni⁴, Eran Segal^{2,3§} & Eran Elinav^{1§}

In tackling the obesity pandemic, considerable efforts are devoted to the development of effective weight reduction strategies, yet many dieting individuals fail to maintain a long-term weight reduction, and instead undergo excessive weight regain cycles. The mechanisms driving recurrent post-dieting obesity remain largely elusive. Here we identify an intestinal microbiome signature that persists after successful dieting of obese mice and contributes to faster weight regain and metabolic aberrations upon re-exposure to obesity-promoting conditions. Faecal transfer experiments show that the accelerated weight regain phenotype can be transmitted to germ-free mice. We develop a machine-learning algorithm that enables personalized microbiome-based prediction of the extent of post-dieting weight regain. Additionally, we find that the microbiome contributes to diminished post-dieting flavonoid levels and reduced energy expenditure, and demonstrate that flavonoid-based 'post-biotic' intervention ameliorates excessive secondary weight gain. Together, our data highlight a possible microbiome contribution to accelerated post-dieting weight regain, and suggest that microbiome-targeting approaches may help to diagnose and treat this common disorder.

The past century has witnessed an alarming increase in the prevalence of obesity, with over 44% of the adult world population estimated to be overweight, and over 300 million adults suffering from morbid obesity. Obesity is considered a major risk factor for 'metabolic syndrome' and its complications, with consequences for life expectancy, quality of life and healthcare costs¹.

Despite continuous medical and scientific effort, long-term strategies aimed at attenuating or reversing the obesity epidemic have yielded disappointing results. While a plethora of dietary approaches efficiently induce weight reduction, in up to 80% of cases in which weight loss was initially successful, reduced weight is not maintained, and instead is followed by recurrent weight gain and relapsing metabolic complications within 12 months of initial weight reduction that may even exceed the pre-dieting metabolic derangements². Post-dieting weight regain is substantially influenced by non-genetic factors, as exemplified by progressively worsening weight regain in weight-cycling twins as compared to their non-dieting siblings³, and is suggested to be independent of starting weight⁴ and level of exercise⁵. Thus, the mechanisms underlying the weight-cycling-induced obesity phenomenon, commonly referred to as the 'yo-yo effect', remain to be determined^{6,7}.

An emerging factor affecting human metabolic homeostasis and the risk for obesity and its metabolic complications is the intestinal microbiome. Compositional and functional microbiome alterations, termed dysbiosis, have been suggested to contribute to the pathogenesis of obesity in both animal models and humans^{8,9}. Moreover, dietary changes have been demonstrated to be central drivers of microbiome composition and function, and markedly affect the microbiome within days of initiation^{10,11}.

In this study, we used mouse models of weight loss and recurrent obesity to investigate the mechanisms underlying exacerbated

metabolic complications following weight cycling. We find that, in contrast to obesity-associated metabolic derangements that can be efficiently reverted upon dieting, obesity-induced alterations to the microbiome persist over long periods of time and enhance the rate of weight regain upon encounter of a second metabolic challenge. This post-dieting dysbiosis affects intestinal flavonoid levels, which, in turn, may impact host energy expenditure. We further devise a machine-learning algorithm that successfully predicts the personalized propensity for recurrent diet-induced obesity solely on the basis of microbiome composition and demonstrate that faecal microbiome transplantation (FMT) or metabolite-based treatment may ameliorate exacerbated post-dieting weight regain.

Enhanced weight regain after dieting

To study the mechanisms modulating post-dieting weight regain, we used a mouse model of recurrent obesity, in which mice were exposed to cycles of a high-fat diet (HFD), interleaved by normal chow (NC) consumption (cycHFD, Fig. 1a). Consequently, in these mice, weight gain and metabolic syndrome developed during primary exposure to HFD, followed by recuperation and weight reduction during exposure to NC, and then by re-emergence of weight gain and associated metabolic disturbances in subsequent HFD-mediated obesity cycles. As controls, we used mice continuously fed a HFD, mice continuously fed a NC diet, and mice that were exposed to only a single cycle of HFD (primHFD, Fig. 1a). As observed in recurrently dieting humans², a preceding obesity-weight-loss cycle rendered mice susceptible to accelerated secondary weight gain, even after fully returning to baseline weight (Fig. 1b, c and Extended Data Fig. 1a, b). As a result, the net weight gain, that is, the weight induced during identical durations of high-fat feeding, was higher in the weight-cycling group compared to

¹Immunology Department, Weizmann Institute of Science, 76100 Rehovot, Israel. ²Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, 76100 Rehovot, Israel. ³Department of Molecular Cell Biology, Weizmann Institute of Science, 76100 Rehovot, Israel. ⁴Department of Plant and Environmental Sciences, Weizmann Institute of Science, 76100 Rehovot, Israel. ⁵Department of Veterinary Resources, Weizmann Institute of Science, 76100 Rehovot, Israel. ⁶The Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University, Rehovot 76100, Israel. ⁷Research Center for Digestive Tract and Liver Diseases, Tel Aviv Sourasky Medical Center, Sackler Faculty of Medicine, Tel Aviv University, 69978 Tel Aviv, Israel.

⁸Digestive Center, Tel Aviv Sourasky Medical Center, 64239 Tel Aviv, Israel.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

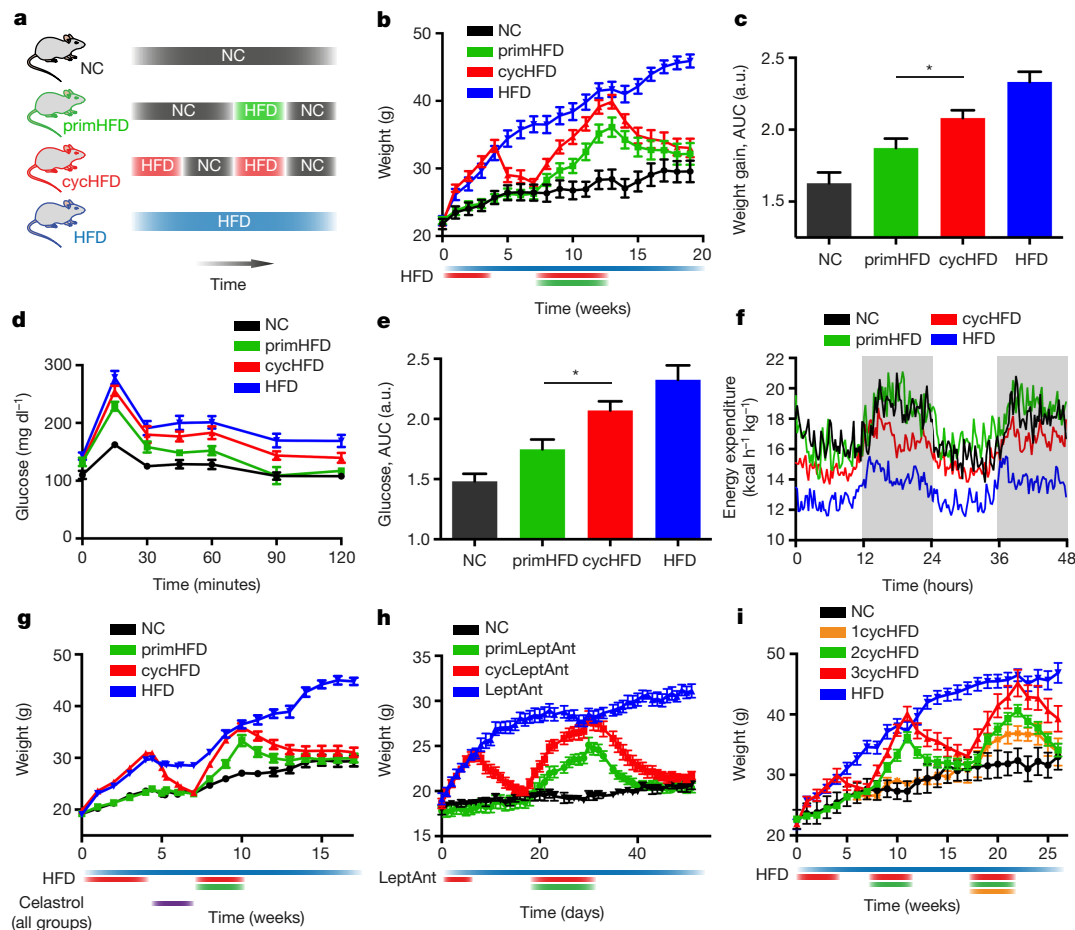


Figure 1 | Enhanced recurrent weight gain after treatment of obesity. **a–c**, Schematic (**a**), weight curves (**b**), and weight regain quantification (**c**) of mice undergoing weight cycling and controls. a.u., arbitrary units. **d–f**, Glucose levels after glucose tolerance test (**d**), glucose quantification (**e**) and energy expenditure over 48 h (**f**) during weight regain (week 10).

mice continuously fed a HFD (Extended Data Fig. 1c). The maximal weight reached by both weight-cycling and continuous HFD groups was comparable, and higher than in mice exposed to a first cycle of HFD (Fig. 1b). Moreover, as compared to mice exposed to a single cycle of HFD, recurrent obesity was characterized by a significant increase in total body fat as determined by MRI (Extended Data Fig. 1d, e), enhanced glucose intolerance (Fig. 1d, e), and elevated serum levels of leptin and low-density lipoprotein (LDL), but not high-density lipoprotein (HDL) (Extended Data Fig. 1f–h). Accelerated weight regain in post-dieting mice was associated with decreased energy expenditure (Fig. 1f and Extended Data Fig. 1i–p), while food intake remained unaffected (Extended Data Fig. 1q).

Similarly, exacerbated metabolic derangements following a weight gain/weight reduction cycle were observed when weight loss was pharmacologically accelerated by celastrol, a quinone methide recently found to induce weight loss¹² (Extended Data Fig. 2a, b). Upon reinstatement of HFD, mice developed significantly exacerbated secondary weight gain (Fig. 1g, Extended Data Fig. 2c–e), as compared to celastrol-treated controls without preceding obesity. Furthermore, to model recurrent obesity induced by hyperphagia rather than dietary composition, we pharmacologically inhibited leptin signalling^{13,14}. Mice administered a leptin antagonist for one week while consuming NC significantly gained weight, and fully returned to normal weight upon cessation of leptin antagonist treatment (Fig. 1h). Upon a second challenge with the leptin antagonist, these mice featured a more pronounced weight regain compared to mice administered the leptin antagonist for the first time (Fig. 1h and Extended Data Fig. 2f, g).

g–i, Recurrent weight gain in mice treated with celastrol (**g**), leptin antagonist (LeptAnt; **h**), or up to three HFD cycles (**i**). Coloured bars below weight curves depict durations of the indicated treatments. Experiments were repeated at least twice. Data are mean \pm s.e.m. * $P < 0.05$ by ANOVA. See Supplementary Tables 5 and 6 for exact n and P values.

When exposed to a third cycle of HFD-induced obesity, weight-cycling mice exhibited a further exacerbation in weight gain (Fig. 1i), obesity (Extended Data Fig. 2h), and dyslipidemia (Extended Data Fig. 2i) as compared to control animals experiencing secondary or primary weight-gain cycles. Together, these experiments suggest that previous obesity–dieting cycles progressively enhance the susceptibility to accelerated weight regain and associated metabolic complications.

Persistence of post-dieting microbiome alterations

Given the above results, we proposed that initial obesity had caused persistent abnormalities that predisposed mice to relapsing metabolic disease upon re-feeding with HFD. We therefore performed metabolic profiling during the first episode of obesity (Extended Data Fig. 2j) and at the ‘nadir’ phase (Extended Data Fig. 3a), that is, when previously obese mice had returned to normal weight that was indistinguishable from that of NC-fed controls. Despite marked metabolic derangements during the primary obesity phase (Extended Data Fig. 2k–o), neither body fat content, serum cholesterol, glucose tolerance, nor serum insulin levels were significantly different between post-dieting mice and their non-cycling controls during the nadir post-obesity phase (Extended Data Fig. 3b–f). Similarly, other hallmarks of obesity, such as oxygen consumption, energy expenditure, physical activity, as well as food and drink intake fully returned to normal baseline levels upon weight loss during the nadir phase (Extended Data Fig. 3g–l and Extended Data Fig. 4a–l).

In contrast, the composition of the intestinal microbiota, which had assumed a dysbiotic state during the primary obesity phase,

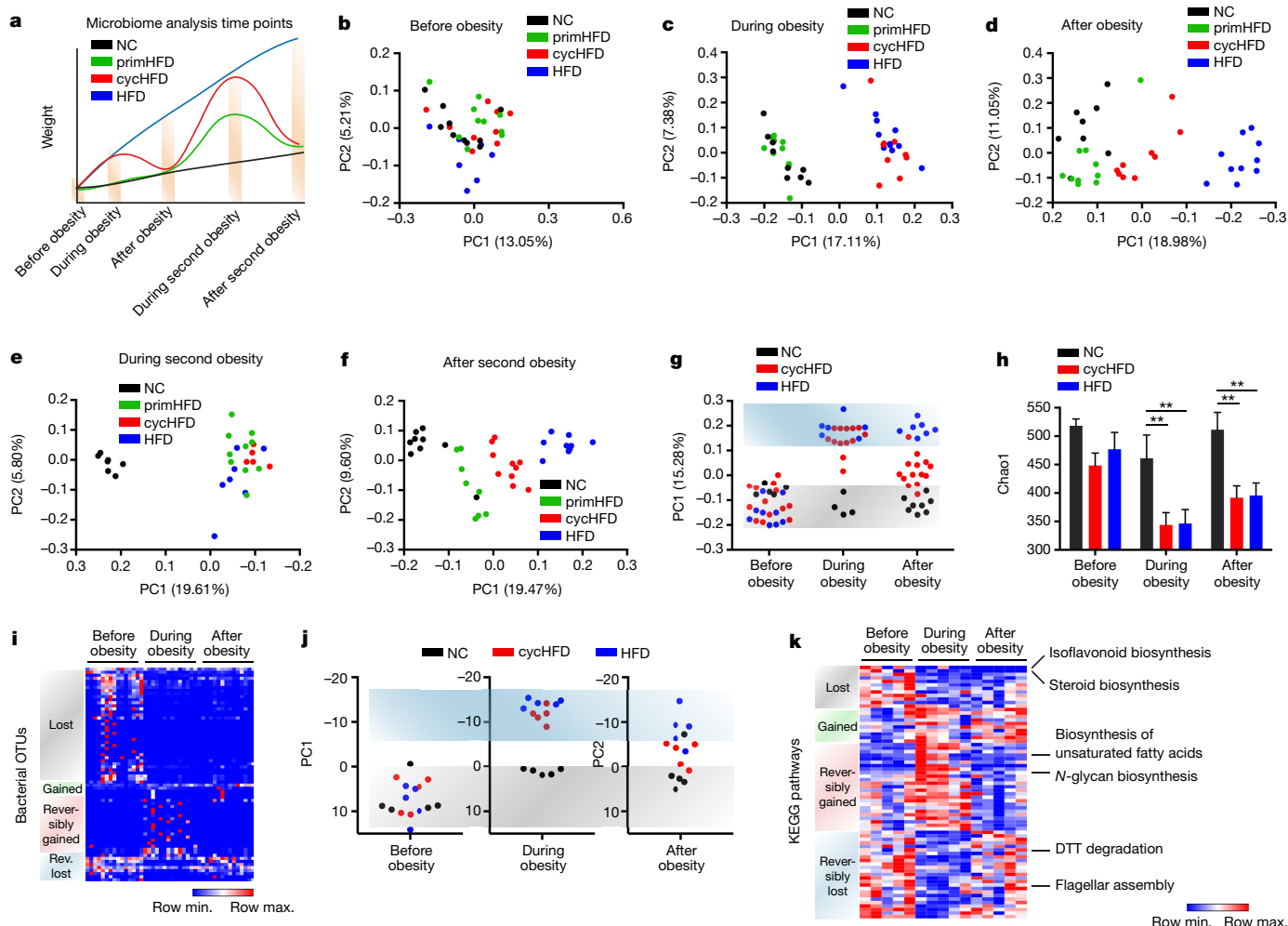


Figure 2 | Persistent microbiome alterations after weight loss.

a, Schematic of sampling times for microbiota analysis. **b–f**, Principal coordinate analyses (PCoA) of unweighted UniFrac distances of microbiota composition at the indicated time points. **g–i**, PCoA (**g**), alpha diversity (**h**), and OTU heatmap (**i**) of weight-cycling mice and controls

did not return to its original composition at the time of post-dieting weight and metabolic normalization (Fig. 2a–d and Extended Data Fig. 5a). Instead, the microbiota assumed an intermediate configuration between the dysbiotic and normal states (Fig. 2d and Extended Data Fig. 5a). A similar microbiome configuration shift was observed after recovery from a second cycle of recurrent obesity (Fig. 2e, f and Extended Data Fig. 5a). We confirmed these findings by targeting an alternative region of the 16S locus for amplicon sequencing (Extended Data Fig. 5b–d). Notably, in addition to the significant alteration in bacterial composition persisting after metabolic normalization (Fig. 2g), bacterial alpha diversity was reduced during the obese state and did not recover upon return to normal weight and metabolic homeostasis (Fig. 2h). To determine the operational taxonomic units (OTUs) that remained altered after dieting, we normalized the OTU abundance of weight cycling mice to age-matched NC controls and classified the OTUs according to their temporal behaviour. Notably, only 45% of all OTUs returned to pre-obesity levels after dieting (Fig. 2i and Extended Data Fig. 5e, f), while obesity-induced effects on the microbiome persisted in multiple bacterial taxa (Extended Data Fig. 5g–l). Similarly, persistent post-obesity microbiota alterations in composition and alpha diversity were noted in mice in which weight loss had been aided with celastrol treatment (Extended Data Fig. 5j–m).

To determine the functional consequences of this incomplete post-dieting microbiota recovery, we performed shotgun metagenomic

before, during, and after obesity. **j**, **k**, Principal component analysis (PCA; **j**) and heatmap (**k**) of bacterial KEGG pathways in weight-cycling mice and controls. Experiments were repeated twice. Data are mean \pm s.e.m. $^{**}P < 0.01$ by ANOVA. See Supplementary Tables 5 and 6 for exact *n* and *P* values.

sequencing and normalized the temporal behaviour of gene abundances to NC controls. We identified 773 bacterial genes whose abundance was altered by HFD and did not return to control levels after dieting (Extended Data Fig. 6a–c). Likewise, microbial functionalities did not fully recover in previously dieting mice, both at the level of gene modules (Extended Data Fig. 6d) and functional pathways (Fig. 2j). Similar to OTUs, reversal of obesity led only to a partial recovery of microbial functions (Extended Data Fig. 6e–h), with multiple obesity-induced microbiome aberrations persisting during weight loss (Fig. 2k). Abundances of genes from multiple metabolic pathways, including isoflavonoid and steroid biosynthesis, were reduced during high-fat feeding and did not recover upon dieting (Extended Data Fig. 6i, j). Collectively, these data indicate that reversal of obesity by dieting results in a microbiome configuration that remains altered, as compared to control mice without prior obesity, even when a state of metabolic normalization is reached.

The post-dieting microbiome contributes to weight regain

We next sought to investigate whether the persistent post-obesity microbiome signature was causally involved in the metabolic complications associated with recurrent weight gain. To this end, we treated mice with broad-spectrum antibiotics during the post-obesity weight loss period (Fig. 3a). Expectedly, antibiotic treatment during dieting

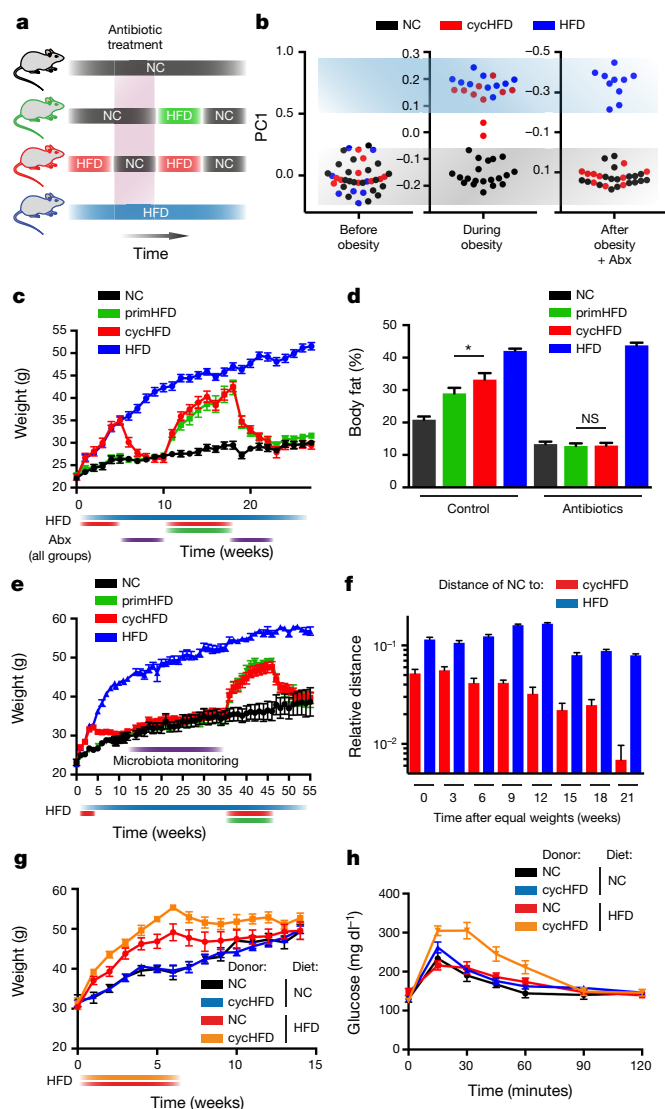


Figure 3 | Post-dieting microbiome alterations drive exacerbated weight regain. **a**, Schematic of microbiota equilibration by antibiotics during weight loss. **b–d**, PCoA of faecal microbiota (**b**), weight curve (**c**) and body fat content (**d**) during weight regain of weight-cycling mice undergoing antibiotic (Abx) treatment during weight loss. **e**, Weight curve of mice monitored for microbiota equilibration after dieting before induction of secondary obesity. **f**, Microbial dissimilarity expressed as differential UniFrac distance over time between cycHFD and HFD mice compared to NC controls. **g, h**, Weight curve (**g**) and glucose levels after glucose tolerance test (**h**) in germ-free recipients one week after microbiota transfer from cycHFD and NC mice. Experiments were repeated twice. Data are mean \pm s.e.m. * $P < 0.05$ by ANOVA; NS, not significant. See Supplementary Tables 5 and 6 for exact n and P values.

abolished the post-obesity microbiome signature and equilibrated the microbiota composition and alpha diversity between previously obese mice and controls, while the microbiota of untreated mice maintained an intermediate configuration (Fig. 3b and Extended Data Fig. 7a, b). Remarkably, antibiotic treatment also abrogated the exacerbation of metabolic derangements upon re-exposure to a HFD, including weight gain, body fat content and glucose intolerance, as compared to untreated weight-cycling mice (Fig. 3c, d and Extended Data Fig. 7c–e).

We next determined the time required for spontaneous reversal of the persistent post-dieting microbiome alterations. To this end, we monitored the microbiota composition every three weeks upon return of previously obese mice to phenotypic normality (Fig. 3e). Notably,

spontaneous reversion of the post-cycling microbiota composition back to NC configuration was achieved only 21 weeks after the completion of successful dieting, that is, a time period more than five times longer than the initial weight gain or dieting period (Fig. 3f and Extended Data Fig. 7f–h). The return to compositional normality in the post-obesity microbiota was associated with gradual acquisition or loss of bacterial taxa (Extended Data Fig. 7i–k). Importantly, following the spontaneous microbiota equilibration, re-exposure to a HFD resulted in an indistinguishable weight gain of formerly obese mice compared to the primary weight gain of HFD-fed mice (Fig. 3e and Extended Data Fig. 7l).

Additionally, we performed faecal transfer experiments, in which the microbiota from previously obese mice (cycHFD) and from phenotypically identical controls (NC) were transferred to germ-free mice, which were subsequently fed either NC or a HFD (Extended Data Fig. 8a). The compositional differences between the microbiota from formerly weight-cycling mice and controls persisted in germ-free recipients (Extended Data Fig. 8b–d). Notably, under NC conditions, naive and post-cycling microbiome-transplanted germ-free mice featured similar weight and glucose tolerance (Fig. 3g, h and Extended Data Fig. 8e, f), suggesting that the post-obesity microbiome did not feature intrinsic obesogenic properties. In contrast, when fed a HFD, recipients of post-weight cycling microbiota exhibited significantly enhanced weight gain and glucose intolerance as early as one week after faecal transplantation, as compared to recipients of microbiome from NC-consuming mice (Fig. 3g, h and Extended Data Fig. 8e, f). Thus, enhanced metabolic derangements in cycling microbiome-transplanted HFD-fed germ-free mice developed even in the absence of previous bouts of obesity in recipient mice, indicating that the post-dieting microbiome configuration coupled with a secondary obesogenic challenge suffices to induce an enhanced metabolic phenotype. Together, these data suggest that the post-dieting microbiota contributes to the susceptibility to develop aggravated metabolic complications upon re-exposure to obesity-inducing conditions.

Microbiota composition predicts weight regain

Given the above causal connection between microbiome configuration and post-dieting weight regain, we asked whether the extent of recurrent weight gain could be computationally predicted for each individual mouse based on its microbiota composition at the post-dieting nadir period. We therefore profiled the microbiota composition of 25 mice that had undergone post-obesity dieting until metabolic normality and 25 weight-matched NC controls (Fig. 4a). We first devised a machine-learning algorithm, based solely on the microbiota composition, aimed at predicting a history of obesity or lack thereof (Fig. 4a, see Methods). Notably, the derived random forest classifier predicted obesity history nearly perfectly ($AUC = 0.96$, Fig. 4b).

We then attempted to predict the exact extent of weight regain of each mouse upon secondary exposure to HFD, based on the mouse's post-dieting microbiome configuration. Prediction solely on the basis of obesity history (that is, without any machine learning model employed) yielded a low prediction accuracy ($R = 0.21$, Extended Data Fig. 8g). By contrast, a 16S rDNA-based prediction using a leave-one-out cross-validation scheme performed significantly better ($R = 0.58$, Extended Data Fig. 8h). Notably, a two-step microbiome-based algorithm that first predicts obesity history and then predicts weight regain on the basis of the predicted history, achieved a highly accurate prediction of the extent of weight regain across individual mice ($R = 0.72$, Fig. 4c).

To determine which features of the microbiota contribute to the algorithm's ability to predict the degree of weight gain, we ranked all OTUs detected in the faecal microbiomes according to feature importance for prediction (Fig. 4d). Notably, a total of 189 OTUs contributed to the algorithm's predictions and the magnitude of their contributions displayed a continuum with no OTU standing out as a major contributor (Fig. 4e). This suggests that the composition of the commensal

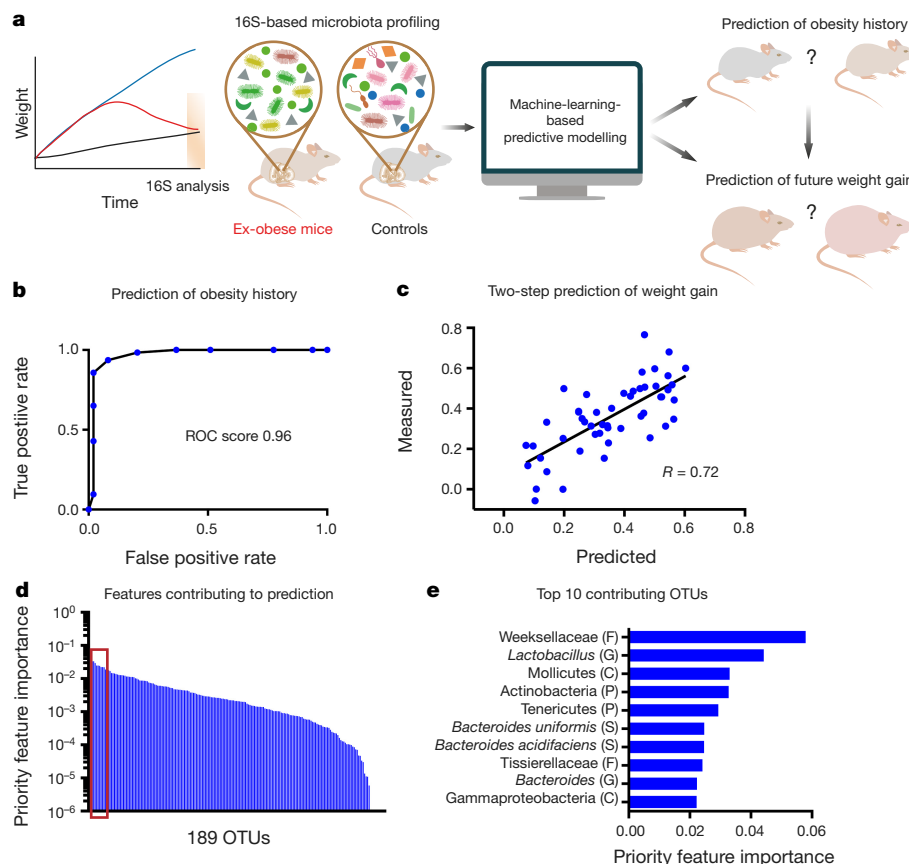


Figure 4 | Accurate prediction of post-dieting weight regain by microbiota features. **a**, Schematic of microbiota-based prediction of weight-gain history and weight regain upon HFD feeding. **b**, **c**, Prediction of prior obesity (**b**) and weight regain (**c**) based on 16S data. **d**, Ranked list

of OTUs contributing to algorithm-based weight prediction. **e**, List of ten most important OTUs from **d**. Letters in brackets indicate taxonomic level (F, family; G, genus; C, class; P, phylum; S, species). See Supplementary Table 5 for exact *n* values.

bacteria as a whole, rather than a small subset of species, drives the post-dieting microbiome alterations that contribute to the susceptibility to relapsing obesity. Together, these results indicate that the microbiota configuration may be used to predict the history of HFD-induced obesity in mice as well as the extent of personalized weight regain that occurs upon recurrence of similar obesity-inducing conditions.

Metabolites contribute to post-dieting weight regain

We next determined whether microbiome modulation during the post-weight-cycling period could ameliorate the extent of secondary weight gain and its metabolic complications. To this end, we performed daily faecal microbiome transplantation (FMT) for 4 weeks, using 'naïve' or post-dieting donor microbiomes transferred into colonized weight-cycling mice during the nadir post-dieting period (Fig. 5a). Upon transplantation, compositional differences between the microbiota from formerly weight-cycling mice and controls persisted in the corresponding FMT recipients (Extended Data Fig. 8i–k). Notably, recipients of a non-cycling 'healthy' microbiome during the nadir post-obese period exhibited an ameliorated secondary weight gain (Fig. 5b and Extended Data Fig. 8l, m), reduced glucose intolerance (Fig. 5c and Extended Data Fig. 8n), decreased body fat (Fig. 5d, e), and increased lean mass (Extended Data Fig. 8o) as compared to mice undergoing a control FMT with a post-cycling microbiome. These results indicate that restoration of normal microbiota function after dieting may prevent exacerbation of metabolic derangements upon weight regain.

Given the effectiveness of FMT, we sought to gain further insight into how microbiota replenishment ameliorates the propensity for recurrent obesity after weight loss. To this end, we longitudinally compared the faecal metabolomics profiles between mice undergoing HFD-induced weight-cycling and control NC-fed mice. We normalized the metabolite

levels in weight-cycling mice to those of age-matched NC controls at each time point and then classified each metabolite according to temporal patterns. As expected, HFD induced major alterations in the faecal metabolome, which were partially reversed upon subsequent weight loss (Fig. 5f, g and Extended Data Fig. 9a–d). However, in nearly half of all metabolites altered by HFD, including several bile acids, the obesity-induced changes persisted after return to phenotypic normality (Fig. 5f, g and Extended Data Fig. 9e–h).

Among the metabolites most significantly depleted by HFD whose levels did not recover upon regain of metabolic health were the dietary flavonoids apigenin and naringenin (Fig. 5f–i). Both compounds remained suppressed for as long as 15 weeks after weight normalization (Fig. 5h, i). Flavonoids are commonly ingested diet-derived compounds that are metabolized by the intestinal microbiota¹⁵. The microbiome contribution to intestinal flavonoid levels was evident from the elevated levels of apigenin and naringenin in antibiotics-treated or germ-free mice (Fig. 5j). We therefore hypothesized that a combination of dietary flavonoid availability and microbiome-mediated flavonoid-degrading capacity may contribute to the total intestinal flavonoid pool. To this end, we followed the kinetics of flavonoids, the flavonoid-biosynthetic enzyme chalcone synthase and the flavonoid-degrading enzyme flavanone 4-reductase (Extended Data Fig. 9i and Supplementary Table 1), over the course of an obesity/weight loss/weight regain cycle (Extended Data Fig. 10a).

During primary obesity, low intestinal flavonoid levels (Fig. 5h, i) were contributed to by low flavonoid availability in the HFD (Extended Data Fig. 10b) coupled to a microbiome shift towards a flavonoid-degrading configuration, as evident by the increase in flavanone 4-reductase levels, and no similar increase in the level of chalcone synthase (Extended Data Fig. 10c, d). During induction of weight loss by

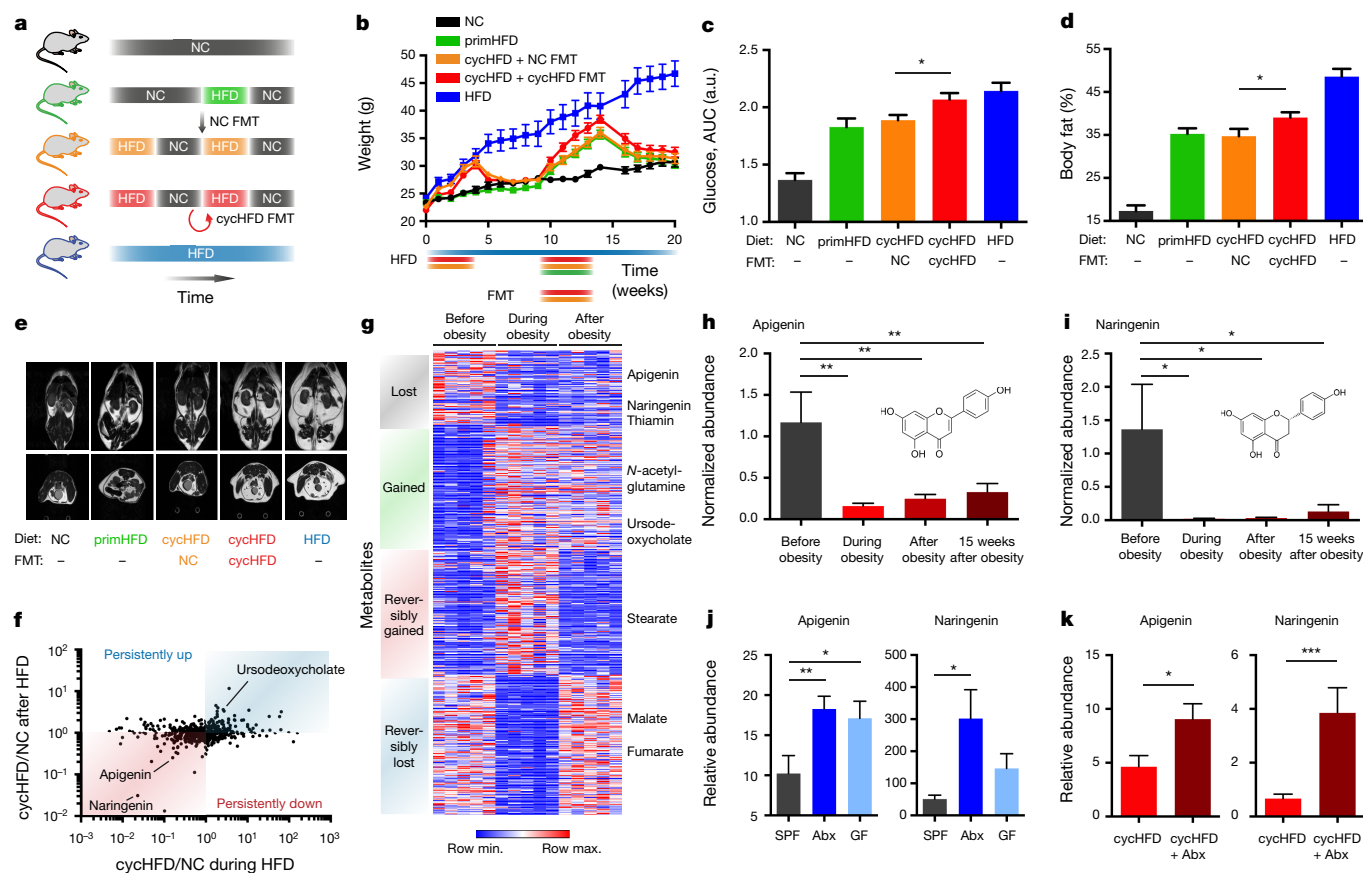


Figure 5 | Microbiome modulation ameliorates post-dieting weight regain. **a–e**, Schematic (**a**), weight curve (**b**), glucose level quantification after glucose tolerance test (**c**), body fat content (**d**), and representative MRI scans (**e**) during weight regain of weight-cycling mice undergoing faecal microbiota transplantation (FMT). **f, g**, Relative abundances (**f**) and heatmap (**g**) of intestinal metabolites before, during, and after obesity. **h–k**, Normalized intestinal abundances of apigenin and naringenin before,

during, and after obesity, as well as 15 weeks after successful dieting (**h, i**), in antibiotic-treated (Abx), germ-free (GF), and specific-pathogen-free (SPF) mice (**j**), and in mice treated with antibiotics during weight loss (**k**). Experiments were repeated twice. Data are mean \pm s.e.m. * P < 0.05, ** P < 0.01, *** P < 0.001 by ANOVA (**c–j**) or Mann–Whitney U -test (**k**). See Supplementary Tables 5 and 6 for exact n and P values.

reversion to a NC diet, dietary flavonoid availability returned to its normally high level (Extended Data Fig. 10b) yet intestinal flavonoid levels remained persistently low, including those of naringenin-derived eriodictyol (Fig. 5h, i and Extended Data Fig. 10e). At this nadir phase, the flavonoid-degrading microbiome contributed to the persistently low flavonoid levels, as suggested by elevated levels of flavanone 4-reductase (Extended Data Fig. 10c) and by the effect of antibiotic treatment during the weight-reduction phase, which diminished the levels of flavanone 4-reductase (Extended Data Fig. 10f) and normalized flavonoid levels (Fig. 5k). Upon acute secondary induction of obesity (Extended Data Fig. 10g), the combination of low dietary flavonoid availability (Extended Data Fig. 10b), coupled with the long-standing presence of a flavonoid-degrading microbiome and associated low flavonoid levels, probably contributed to reduced flavonoid levels in weight-cycling mice as compared to controls undergoing primary weight gain (Extended Data Fig. 10h, i).

Apigenin and naringenin have been reported to affect food intake, adipocyte differentiation, and lipid metabolism^{16–18}. We therefore hypothesized that, similarly to FMT-induced restoration of a naive flavonoid-degrading microbiome in weight-cycling mice, direct flavonoid replenishment in these mice may ameliorate the exacerbated relapsing obesity phenotype. Indeed, oral daily administration of both flavonoids to post-dieting mice during the post dieting nadir and secondary weight regain period (Fig. 6a) resulted in normalization of intestinal apigenin and naringenin levels to control levels (Extended Data Fig. 10j), with no effect noted on microbiome composition (Extended Data Fig. 10k). Similar to FMT, combined flavonoid

treatment ameliorated the rate of secondary weight regain (Fig. 6b, c and Extended Data Fig. 10l, m). Together, these results suggest that low apigenin and naringenin levels in post-dieting mice contribute to an exacerbated weight regain, while their therapeutic replenishment ameliorates this susceptibility.

Flavonoids modulate weight regain and UCP1 expression

To investigate possible mechanisms by which apigenin and naringenin ameliorate recurrent post-dieting obesity, we measured metabolic and behavioural parameters in flavonoid-administered weight-cycling mice, as compared to untreated weight-cycling mice. Notably, weight-adjusted energy expenditure was markedly reduced in weight-cycling mice (Fig. 6d–f and Extended Data Fig. 11a–f), but was normalized upon flavonoid administration (Fig. 6d–f and Extended Data Fig. 11a–f), suggesting that apigenin and naringenin might impact host energy expenditure. Of note, a similar effect of flavonoids on weight management was observed in previous studies¹⁹, while the link between flavonoid supplementation and enhanced energy expenditure was reached only upon normalization of energy expenditure to body weight. Other metabolic parameters were not affected by flavonoid treatment (Extended Data Fig. 11g, h and Extended Data Fig. 12a–f). Similarly, mice treated with antibiotics during the nadir period featured enhanced energy expenditure upon re-administration of HFD (Fig. 6g, Extended Data Fig. 12g–l and Extended Data Fig. 13a–c), in line with higher levels of flavonoids (Fig. 5k) and amelioration of the exacerbated weight regain (Fig. 3c), while not altering other metabolic parameters during recurrent weight gain (Extended Data Fig. 13d–l).

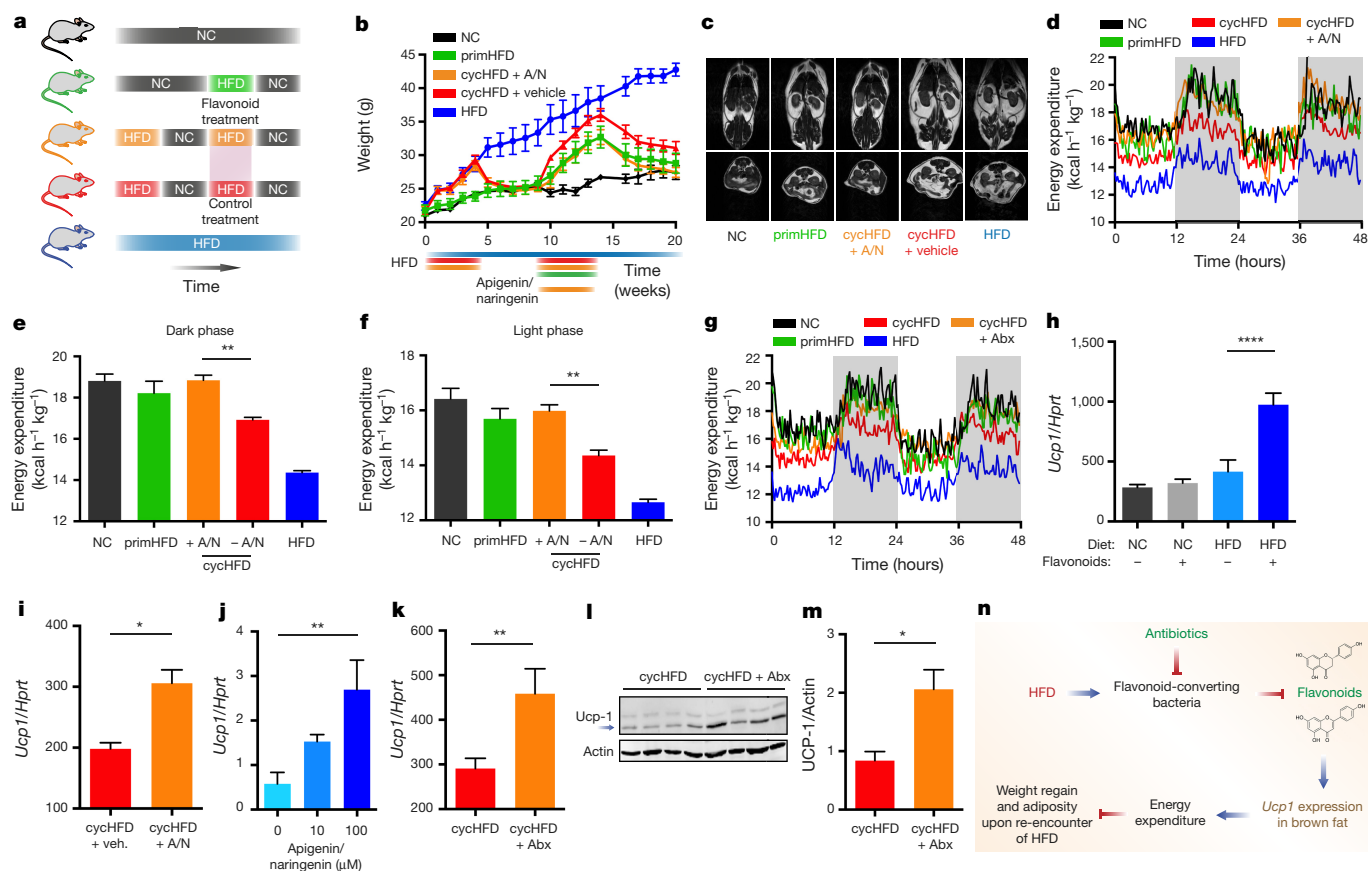


Figure 6 | Metabolite modulation ameliorates post-dieting weight regain. **a–f**, Schematic (**a**), weight curve (**b**), representative MRI scans (**c**), representative energy expenditure recording (**d**), and quantifications during the dark phase (night) (**e**) and light phase (day) (**f**) of weight-cycling mice with or without supplementation of apigenin and naringenin (A/N) during weight regain. **g**, Energy expenditure recording during weight regain of weight-cycling mice with or without antibiotic treatment before weight regain. **h–m**, Brown adipose tissue (BAT) *Ucp1* transcript (**h–k**) or UCP1 protein (**l, m**) levels in mice on HFD or NC receiving A/N

Finally, we pursued possible mechanisms by which flavonoids may participate in the regulation of host energy expenditure. Since brown adipose tissue (BAT) is a major regulator of thermogenesis in mammals, and since other members of the flavonoid family have been previously associated with the induction of the major thermogenic factor uncoupling protein 1 (UCP1)^{20–22}, we analysed *Ucp1* expression in mice fed NC or HFD and orally administered with apigenin and naringenin. As early as two weeks after the start of flavonoid treatment, *Ucp1* transcript levels were significantly elevated in the BAT of mice fed a HFD, but not in those fed NC (Fig. 6h). Likewise, *Ucp1* expression was elevated in weight-cycling mice upon apigenin and naringenin supplementation during the HFD-induced weight-regain period (Fig. 6i). *Ucp1* was also induced by flavonoids in BAT explants in a concentration-dependent manner (Fig. 6j), suggesting that there exists a direct effect of apigenin and naringenin on the modulation of gene expression in BAT. Given that antibiotic treatment elevated intestinal flavonoid levels (Fig. 5k) and host energy expenditure (Fig. 6g and Extended Data Fig. 12h, l), we determined the levels of UCP1 in the BAT of antibiotic-treated weight-cycling mice. Indeed, we found both transcript and protein levels of UCP1 to be elevated in the group receiving antibiotics (Fig. 6k–m), providing a potential mechanistic explanation for recent observations of *Ucp1* induction in germ-free mice²³.

Taken together, these associations suggest a model in which HFD promotes the growth of flavonoid-metabolizing bacteria, which in turn decrease the amount of bioavailable flavonoids, thereby negatively

or vehicle by daily gavage for 2 weeks (**h**), in weight-cycling mice with or without A/N treatment during weight regain (**i**), in BAT explants cultured with A/N for 24 h (**j**), and in mice during weight regain after antibiotic treatment during weight loss (**k–m**). **n**, Model of diet-microbiota-energy-expenditure interactions during dieting and weight regain. Experiments were repeated twice. Data are mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$ by ANOVA (**e, f, h, j**) or Mann-Whitney *U*-test (**i, k, m**). See Supplementary Tables 5 and 6 for exact *n* and *P* values.

regulating UCP1-driven energy expenditure and promoting exaggerated recurrent weight gain (Fig. 6n). Full validation of this model merits future studies investigating whether the flavonoid effect on BAT *Ucp1* expression directly drives enhanced energy expenditure.

Discussion

In this study, we describe the persistence of an altered microbiome configuration following cycles of obesity and dieting, which contributes to enhanced metabolic derangements upon weight regain, through metabolite-induced effects on host metabolism. We hypothesize that diet-induced microbiome persistence may have evolved to act as a ‘buffer’ contributing to the stability of metabolic homeostasis over prolonged periods of time, by preventing overly fluctuating metabolic responses to incidental nutritional or environmental signals. However, in contexts of erratic changes in host physiology, such as cycling weight gain and dieting, this microbiome persistence may predispose the host to exaggerated metabolic consequences in ensuing weight-gain cycles. Similar contexts of microbiome persistence include its hysteresis-like behaviour of reduced reversibility during recurrent dietary changes²⁴, or following low-fibre feeding²⁵.

Our results highlight two potentially interdependent yet inherently different microbiota effects on weight and metabolism. First, the microbiota from obese donors induces weight gain in faecal transplanted mice even when recipient mice are maintained on a normal chow diet^{8,9}. This dominant obesogenic property is lost upon remission of

obesity. Second, the persistent post-dieting microbiota influences the degree of relapsing obesity following weight cycling, but only upon encounter of a 'second hit' that gives rise to recurrent weight gain. While we suggest one mechanism for the common failure of formerly obese individuals to maintain long-term reduced weight after dieting, the reasons for this failure are probably complex and include contributions from a multitude of behavioural, genetic, environmental, and metabolic factors²⁶. The findings described here suggest that the remission of metabolic derangements after treatment of obesity precedes the remission of dysbiosis, with the time window of post-obesity microbiome persistence marking the susceptibility phase for accelerated recurrent obesity. Corroboration of these findings in dieting humans as well as additional variables not reflected in mouse models merit further prospective human studies.

Finally, our study provides an example for how rational post-biotic metabolite therapy could serve as a potential means of modulating physiological function downstream of the microbiota. As such, we found that obesity-induced loss of the flavonoids apigenin and naringenin enhances the susceptibility for accelerated weight regain, potentially through impairment of energy expenditure, while replenishment of these metabolites ameliorated these metabolic abnormalities. Future studies are warranted to examine the potential clinical use of flavonoids, as well as modulation of other bioactive metabolites such as bile acids that we found to be persistently elevated after dieting, as novel therapeutics in the quest for effective long-term weight management solutions.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 February; accepted 18 November 2016.

Published online 24 November 2016.

- Stevens, J., Oakkar, E. E., Cui, Z., Cai, J. & Truesdale, K. P. US adults recommended for weight reduction by 1998 and 2013 obesity guidelines, NHANES 2007–2012. *Obesity (Silver Spring)* **23**, 527–531 (2015).
- Anastasiou, C. A., Karfopoulou, E. & Yannakoulia, M. Weight regaining: from statistics and behaviors to physiology and metabolism. *Metabolism* **64**, 1395–1407 (2015).
- Pietiläinen, K. H., Saarni, S. E., Kaprio, J. & Rissanen, A. Does dieting make you fat? A twin study. *Int. J. Obes.* **36**, 456–464 (2012).
- Neumark-Sztainer, D. et al. Obesity, disordered eating, and eating disorders in a longitudinal study of adolescents: how do dieters fare 5 years later? *J. Am. Diet. Assoc.* **106**, 559–568 (2006).
- Saarni, S. E., Rissanen, A., Sarna, S., Koskenvuo, M. & Kaprio, J. Weight cycling of athletes and subsequent weight gain in middleage. *Int. J. Obes.* **30**, 1639–1644 (2006).
- Dulloo, A. G. & Montani, J. P. Pathways from dieting to weight regain, to obesity and to the metabolic syndrome: an overview. *Obes. Rev.* **16** (Suppl 1), 1–6 (2015).
- Mehta, T., Smith, D. L., Jr, Muhammad, J. & Casazza, K. Impact of weight cycling on risk of morbidity and mortality. *Obes. Rev.* **15**, 870–881 (2014).
- Turnbaugh, P. J. et al. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
- Ridaura, V. K. et al. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341**, 1241214 (2013).
- Korem, T. et al. Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* **349**, 1101–1106 (2015).
- David, L. A. et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
- Liu, J., Lee, J., Salazar Hernandez, M. A., Mazitschek, R. & Ozcan, U. Treatment of obesity with celastrol. *Cell* **161**, 999–1011 (2015).
- Elinav, E. et al. Pegylated leptin antagonist is a potent orexigenic agent: preparation and mechanism of activity. *Endocrinology* **150**, 3083–3091 (2009).
- Shpilman, M. et al. Development and characterization of high affinity leptins and leptin antagonists. *J. Biol. Chem.* **286**, 4429–4442 (2011).
- Braune, A. & Blaut, M. Bacterial species involved in the conversion of dietary flavonoids in the human gut. *Gut Microbes* **7**, 216–234 (2016).
- Myoung, H. J., Kim, G. & Nam, K. W. Apigenin isolated from the seeds of *Perilla frutescens* britton var *crispa* (Benth.) inhibits food intake in C57BL/6J mice. *Arch. Pharm. Res.* **33**, 1741–1746 (2010).
- Guo, X., Liu, J., Cai, S., Wang, O. & Ji, B. Synergistic interactions of apigenin, naringin, quercetin and emodin on inhibition of 3T3-L1 preadipocyte differentiation and pancreas lipase activity. *Obes. Res. Clin. Pract.* **10**, 327–339 (2016).
- Assini, J. M. et al. Naringenin prevents obesity, hepatic steatosis, and glucose intolerance in male mice independent of fibroblast growth factor 21. *Endocrinology* **156**, 2087–2102 (2015).
- Hoek-van den Hil, E. F. et al. Direct comparison of metabolic health effects of the flavonoids quercetin, hesperetin, epicatechin, apigenin and anthocyanins in high-fat-diet-fed mice. *Genes Nutr.* **10**, 469 (2015).
- Goldwasser, J. et al. Transcriptional regulation of human and rat hepatic lipid metabolism by the grapefruit flavonoid naringenin: role of PPAR α , PPAR γ and LXRs. *PLoS One* **5**, e12399 (2010).
- Kudo, N. et al. A single oral administration of theaflavins increases energy expenditure and the expression of metabolic genes. *PLoS One* **10**, e0137809 (2015).
- Choi, J. H. & Yun, J. W. Chrysin induces brown fat-like phenotype and enhances lipid metabolism in 3T3-L1 adipocytes. *Nutrition* **32**, 1002–1010 (2016).
- Suárez-Zamorano, N. et al. Microbiota depletion promotes browning of white adipose tissue and reduces obesity. *Nat. Med.* **21**, 1497–1501 (2015).
- Carmody, R. N. et al. Diet dominates host genotype in shaping the murine gut microbiota. *Cell Host Microbe* **17**, 72–84 (2015).
- Sonnenburg, E. D. et al. Diet-induced extinctions in the gut microbiota compound over generations. *Nature* **529**, 212–215 (2016).
- Montani, J. P., Schutz, Y. & Dulloo, A. G. Dieting and weight cycling as risk factors for cardiometabolic diseases: who is really at risk? *Obes. Rev.* **16** (Suppl 1), 7–18 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the members of the Elinav and Segal laboratories for discussions and C. Bar-Nathan for germ-free mouse work. C.A.T. received a Boehringer Ingelheim Fonds PhD Fellowship. D.R. received a Levi Eshkol PhD Scholarship for Personalized Medicine by the Israeli Ministry of Science. Y.K. is the incumbent of the Sarah and Rolando Uziel Research Associate Chair. E.S. is supported by the Crown Human Genome Center; the Else Kroener Fresenius Foundation; D. L. Schwarz; J. N. Halpern; L. Steinberg; and grants funded by the European Research Council, the National Institute of Health, and the Israel Science Foundation. E.E. is supported by Y. and R. Ungar; the Abisch Frenkel Foundation for the Promotion of Life Sciences; the Gurwin Family Fund for Scientific Research; the Leona M. and Harry B. Helmsley Charitable Trust; the Crown Endowment Fund for Immunological Research; the estate of J. Gitlitz; the estate of L. Hershkovich; the Benozio Endowment Fund for the Advancement of Science; the Adels Foundation; J. L. and V. Schwartz; A. and G. Markovitz; A. and C. Adelson; the French National Center for Scientific Research (CNRS); D. L. Schwarz; the V. R. Schwartz Research Fellow Chair; L. Steinberg; J. N. Halpern; A. Edelheit; and by grants funded by the European Research Council; a Marie Curie Integration grant; the German-Israeli Foundation for Scientific Research and Development; the Israel Science Foundation; the Minerva Foundation; the Rising Tide Foundation; the Helmholtz Foundation; and the European Foundation for the Study of Diabetes. E.E. is the incumbent of the Rina Gudinski Career Development Chair and a senior fellow of the Canadian Institute for Advanced Research (CIFAR).

Author Contributions C.A.T. conceived and led the project, designed and performed experiments, and wrote the manuscript. S.I. designed and performed experiments, and wrote the manuscript. D.R. developed and performed bioinformatics methods and analysis, and wrote the manuscript. M.M., M.L., C.M., L.D., S.B. and S.R. performed experiments. S.M., M.D.-B., Y.K. and I.B. performed flavonoid measurements, next-generation sequencing, metabolic measurements, and MRI, respectively. A.G. provided essential tools. A.H., H.S., Z.H. and A.A. provided insights and supervised parts of the experimental work. E.S. and E.E. conceived and directed the project, designed experiments, supervised the participants, and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.S. (eran.segal@weizmann.ac.il) or E.E. (eran.elinav@weizmann.ac.il).

Reviewer Information Nature thanks C. Nagler and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Mice. C57Bl/6 mice were purchased from Harlan and allowed to acclimatize to the animal facility environment for 2 weeks before being used for experimentation. In each experiment, all mice were littermates born and raised in the same vivarium and obtained through a single delivery. All mice were maintained on a strict 12-h light–dark cycle (lights turned on at 6 am and turned off at 6 pm) and were housed in cages containing a maximum of five animals. Numbers of animals were chosen to ensure that a minimum of two distinct cages was used per experimental group in each experiment. No statistical methods were used to predetermine sample size. Mice were taken out of experiments when wounded as a result of fighting among cage-mates. Weights were always measured at the same circadian time throughout each experiment. Other than weight and glucose measurements, investigators were blinded with regard to experimental groups. Outbred Swiss Webster germ-free mice were born in the Weizmann Institute germ-free facility and routinely monitored for sterility. For faecal transplantation experiments, 100 mg of stool was resuspended in 1 ml of PBS under anaerobic conditions, homogenized, and filtered through a 70 µm strainer. Recipient mice were gavaged with 200 µl of the filtrate.

All experiments involving weight cycling employed the following experimental paradigm:

1) Before dietary interventions, mice were randomized to ensure that no incidental pre-diet differences in body weight, body fat content, or microbiome composition existed between the different groups (Supplementary Fig. 1).

2) Initial weight gain for 4 weeks ('during obesity' time point, 4 weeks), see also Supplementary Figs 2 and 3.

3) Weight loss until lean control levels are reached (end-point criterion for weight loss: no statistically significant weight difference between cycling group and lean controls; 'after obesity' time point).

4) Weight regain until obese control levels are reached (end-point criterion for weight gain: no statistically significant weight difference between cycling group and obese controls; 'second obesity' time point).

In all experiments, age-matched male mice were used. In Fig. 1h, female mice were used. Mice were 8 weeks of age at the beginning of experiments. For antibiotic treatment, mice were given a combination of vancomycin (0.5 g l^{-1}), ampicillin (1 g l^{-1}), neomycin (1 g l^{-1}), and metronidazole (1 g l^{-1}) in their drinking water²⁷. Mice were carefully monitored for signs of dehydration upon antibiotic administration. All antibiotics were obtained from Sigma Aldrich and given for the time periods indicated in each figure. For flavonoid treatments, apigenin and naringenin (obtained from Sigma Aldrich) were dissolved in DMSO and administered daily by oral gavage at a concentration of 80 mg kg^{-1} . Controls received vehicle gavages. Celastrol was administered daily by intraperitoneal injection of $100 \mu\text{g kg}^{-1}$ as previously described¹². Leptin antagonist was administered daily by intraperitoneal injection of 25 mg kg^{-1} as previously described¹³.

Rodent diets are detailed in the Supplementary Information (Supplementary Tables 2 and 3). Stool samples were collected fresh and on the basis of individual mice. Fresh stool samples were collected into tubes, immediately snap-frozen in liquid nitrogen upon collection, and stored at -80°C until DNA isolation. All experimental procedures were approved by the local IACUC.

Glucose tolerance test. Mice were fasted for 6 h and subsequently given 200 µl of a 0.2 g ml^{-1} glucose solution (JT Baker) by oral gavage. Blood glucose was determined at 0, 15, 30, 60, 90 and 120 min after glucose challenge (Contour blood glucose meter, Bayer).

Magnetic resonance imaging. Mice were anaesthetized with isoflurane (5% for induction, 1–2% for maintenance) mixed with oxygen (1 l min^{-1}) and delivered through a nasal mask. Once anaesthetized, the animals were placed in a head-holder to assure reproducible positioning inside the magnet. Respiration rate was monitored and kept throughout the experimental period around 60–80 breaths per minute. MRI experiments were performed on 9.4 Tesla BioSpec Magnet 94/20 USR system (Bruker) equipped with gradient coils system capable of producing pulse gradient of up to 40 gauss cm^{-1} in each of the three directions. All MR images were acquired with a quadrature resonator coil (Bruker). The MRI protocol included two sets of coronal and axial multi-slices T2-weighted MR images. The T2-weighted images were acquired using the multi-slice RARE sequence (TR = 2,500 ms, TE = 35 ms, RARE factor = 8), with matrix size being 256×256 , four averages, corresponding to an image acquisition time of 160 s per set. The first set was used to acquire 21 axial slices with 1-mm slice thickness (no gap). The field of view was selected with $4.2 \times 4.2 \text{ cm}^2$. The second set was used to acquire 17 coronal slices with 1-mm slice thickness (no gap). The field of view was selected with $7.0 \times 5.0 \text{ cm}^2$. Total fat and lean mass of mice were quantified by EchoMRI-100 (Echo Medical Systems).

Metabolic measurements. Food intake and locomotor activity were measured using the PhenoMaster system (TSE-Systems), which consists of a combination of sensitive feeding sensors for automated measurement and a photobeam-based

activity monitoring system detects and records ambulatory movements, including rearing and climbing, in each cage. All parameters were measured continuously and simultaneously. Mice were trained singly housed in identical cages before data acquisition.

Triglycerides, total cholesterol and high-density lipoprotein (HDL) levels were measured in mouse serum by SpotChem EZ Chemistry Analyzer (Arkray). LDL levels were calculated using the Friedewald formula.

Concentrations of leptin (Mouse Leptin DUO set, R&D Systems) and insulin (Ultra-sensitive mouse insulin ELISA kit, Crystal Chem) in the serum were measured using ELISA according to the manufacturer's instructions.

Taxonomic microbiota analysis. Frozen faecal samples were processed for DNA isolation using the MoBio PowerSoil kit according to the manufacturer's instructions. 1 ng of purified faecal DNA was used for PCR amplification. Amplicons spanning the variable region 1/2 (V1/2) of the 16S rRNA gene were generated by using the following barcoded primers: forward, 5'-XXXXXXXXXAGAGTTTGATCCTGGCTCAG-3'; reverse, 5'-TGCTGCCTCCCGTAGGAGT-3', where X represents a barcode base. Amplicons spanning the variable region 3/4 (V3/4) of the 16S rRNA gene (Fig. 2h and Extended Data Fig. 5b–d) were generated by using the following primers: forward, 5'-GTGCCAGCMGCCGCGGTAA-3'; reverse 5'-GGACTACHVGGGTWTCTAAT-3'. The reactions were subsequently pooled and cleaned (PCR clean kit, Promega), and the PCR products were then sequenced on an Illumina MiSeq with 500-bp paired-end reads. The reads were then processed using the QIIME (Quantitative Insights Into Microbial Ecology, <http://www.qiime.org>) analysis pipeline as described^{28–30}. In brief, fasta quality files and a mapping file indicating the barcode sequence corresponding to each sample were used as inputs, reads were split by samples according to the barcode, taxonomical classification was performed using the RDP-classifier, and an OTU table was created. OTU mapping was employed using the Greengenes database. Rarefaction was used to exclude samples with insufficient count of reads per sample. Sequences sharing 97% nucleotide sequence identity in the 16S region were binned into operational taxonomic units (97% ID OTUs). For beta diversity, unweighted UniFrac measurements were plotted according to the two principal coordinates based on >10,000 reads per sample. For microbial distance measurements, unweighted UniFrac distances were compared.

Classification of obesity history. Mouse obesity history was predicted using Random Forest Classification (sklearn 0.15.2) with the features being the relative abundances of 16S OTUs as outputted by QIIME. Classification was made in leave-one-out cross-validation in which each mouse was classified as negative or positive for obesity history.

Prediction of weight regain following HFD diet. Future weight gain of mice was predicted in leave-one-out cross-validation, whereby the weight regain of each mouse upon HFD exposure was predicted using a gradient-boosting regression algorithm using data of all other mice consisting of their 16S rDNA OTU data at the post-dieting nadir period, their obesity history, and their individual weight regain upon HFD exposure. Importantly, each time a mouse was left out and its weight regain was predicted, its obesity history was not given as an input to our algorithm. For each left out mouse, a classifier of obesity history was first learned and used to classify the obesity history of the left out mouse as described above. Then, training data mice with the same obesity history as the left out mouse were taken, and gradient-boosting regression (GBR, sklearn 0.15.2) was applied to learn a model that predicts their weight regain on the HFD diet. Input to this model consists of the 16S OTUs which were used within the GBR algorithm to predict weight regain.

Metagenomic analysis. Metagenomic reads containing Illumina adapters were filtered for exclusion of low-quality reads and trimmed low-quality read edges. Host DNA was detected and excluded by mapping with GEM to the mouse genome with inclusive parameters. Length-normalized RA of genes, obtained by similar mapping with GEM to a reference catalogue³¹, was assigned to KEGG Orthology (KO) entries³², and these were then normalized to a sum of 1. RA of KEGG modules and pathways was calculated by summation. Only samples with >100,000 metagenomics reads were considered for analysis.

Quality control of metagenomic reads and removal of host DNA. We applied Trimmomatic³³ with the following parameters:

ILLUMINACLIP:<TruSeq3 adapters FASTA file>:2:30:10 LEADING:25 TRAILING:25 MINLEN:50. We removed host DNA by mapping to the mouse genome (mm10, downloaded from <https://genome.ucsc.edu>) and removing any mapped reads (see section below).

Mapping of metagenomic sequencing reads. Mapping was performed using the GEM mapper³⁴ with the following parameters:

-q offset-33-gem-quality-threshold 26 -m 0.1 -e 0.1-min-matched-bases 0.8-max-big-indel-length 15 -s 3 -d 'all' -D 1 -v -T 2 -p -E 0.3-max-extendable-matches 'all'-max-extensions-per-match 5

With the addition of the modifier -m set to 0.05 for mapping to the mouse genome. Resulted mappings were retained as long as they had at least 50 matched bases with minimal quality of 26.

Genetic content relative abundance calculation. Reads were mapped to the integrated reference catalogue of the human gut microbiome³¹. For each gene in the catalogue, the fraction of reads mapped to it from each sample was counted and normalized by gene length in kilobases. Reads mapping to more than one location were split so that each location received an equal fraction of the mapped read. Mapped reads were subsequently assigned to KEGG Orthology (KO) entries using the gene annotation table available at <http://meta.genomics.cn/>. Relative gene abundances were then calculated by normalizing the KEGG genes of each sample to sum to 1. To calculate the abundances of KEGG pathways and modules, the relative abundance of genes in each pathway and module was summed.

Non-targeted metabolomics. Caecal samples were collected, immediately frozen in liquid nitrogen and stored at -80°C . Sample preparation and analysis was performed by Metabolon Inc. Samples were prepared using the automated MicroLab STAR system (Hamilton). To remove protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated with methanol. The resulting extract was divided into five fractions: one for analysis by UPLC-MS/MS with positive ion mode electrospray ionization, one for analysis by UPLC-MS/MS with negative ion mode electrospray ionization, one for LC polar platform, one for analysis by GC-MS, and one sample was reserved for backup. Samples were placed briefly on a TurboVap (Zymark) to remove the organic solvent. For LC, the samples were stored overnight under nitrogen before preparation for analysis. For GC, each sample was dried under vacuum overnight before preparation for analysis.

Data extraction and compound identification. Raw data was extracted, peak-identified and QC processed using Metabolon's hardware and software. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities.

Metabolite quantification and data normalization. Peaks were quantified using area-under-the-curve. For studies spanning multiple days, a data normalization step was performed to correct variation resulting from instrument inter-day tuning differences.

Flavonoid measurements. Apigenin and naringenin were measured by Waters TQ MS detector combined with Waters Acquity UPLC system. The chromatographic separation was carried out on a BEH C18 column ($1.7\mu\text{m}$, $2.1 \times 100\text{mm}$, Waters). The solvent flow rate was 0.3 ml min^{-1} . The mobile phase consisted of 0.1% formic acid (FA) in 5% acetonitrile (A) and 0.1% FA in acetonitrile (B) using a gradient program described below. The autosampler was cooled to 12°C and the column heated to 35°C . MS detector (Waters TQ) was equipped with an ESI source used in positive mode (capillary voltage 2.5 kV). The measurement was performed in MRM mode, two MRM traces for each compound (one for quantification, and the second for identification). The cone voltages (V) and collision energies (eV) for each MRM transition, as determined by direct injection, are summarized below. Data were processed with MassLynx software with TargetLynx (version 4.1, Waters).

For sample preparation, 50 mg of stool or 100 mg food were weighed into 2-ml safe-lock Eppendorf tubes. Samples were homogenized using a beadbeater with metal balls. Three-hundred micrograms of 80% methanol in DDW were added to the samples, followed by sonication for 20 min, centrifugation, and filtering through Acrodisc PTFE $0.2\mu\text{m}$ filters (P/N 4552T) into vials.

See Supplementary Table 4 for chromatographic conditions for flavonoid separation.

Gene expression analysis. Tissues were preserved in RNAlater solution (Ambion) and subsequently homogenized in Tri Reagent (Sigma Aldrich). RNA was purified using standard chloroform extraction. Two micrograms of total RNA was used to generate cDNA (HighCapacity cDNA Reverse Transcription kit; Applied Biosystems). Real-time PCR was performed using the following *Ucp1* primers: *Ucp1* forward, 5'-GGCCTCTACGACTCAGTCCA-3'; *Ucp1* reverse, 5'-TAA GCCGGCTGAGATCTTGT-3'. Primers for flavanone 4-reductase and chalcone were tested using <http://insilico.ehu.es/PCR/> and validated using cultures of *Lactococcus lactis* and *Escherichia coli*.

PCR was performed using Kapa Sybr qPCR kit (Kapa Biosystems) on a Viia7 instrument (Applied Biosystems). PCR conditions were 95°C for 20 s, followed by 40 cycles of 95°C for 3 s and 60°C for 30 s. Data were analysed using the $\Delta\Delta\text{C}_t$ method with *Hprt* serving as the reference housekeeping gene. *Hprt* cycles were assured to be insensitive to the experimental conditions.

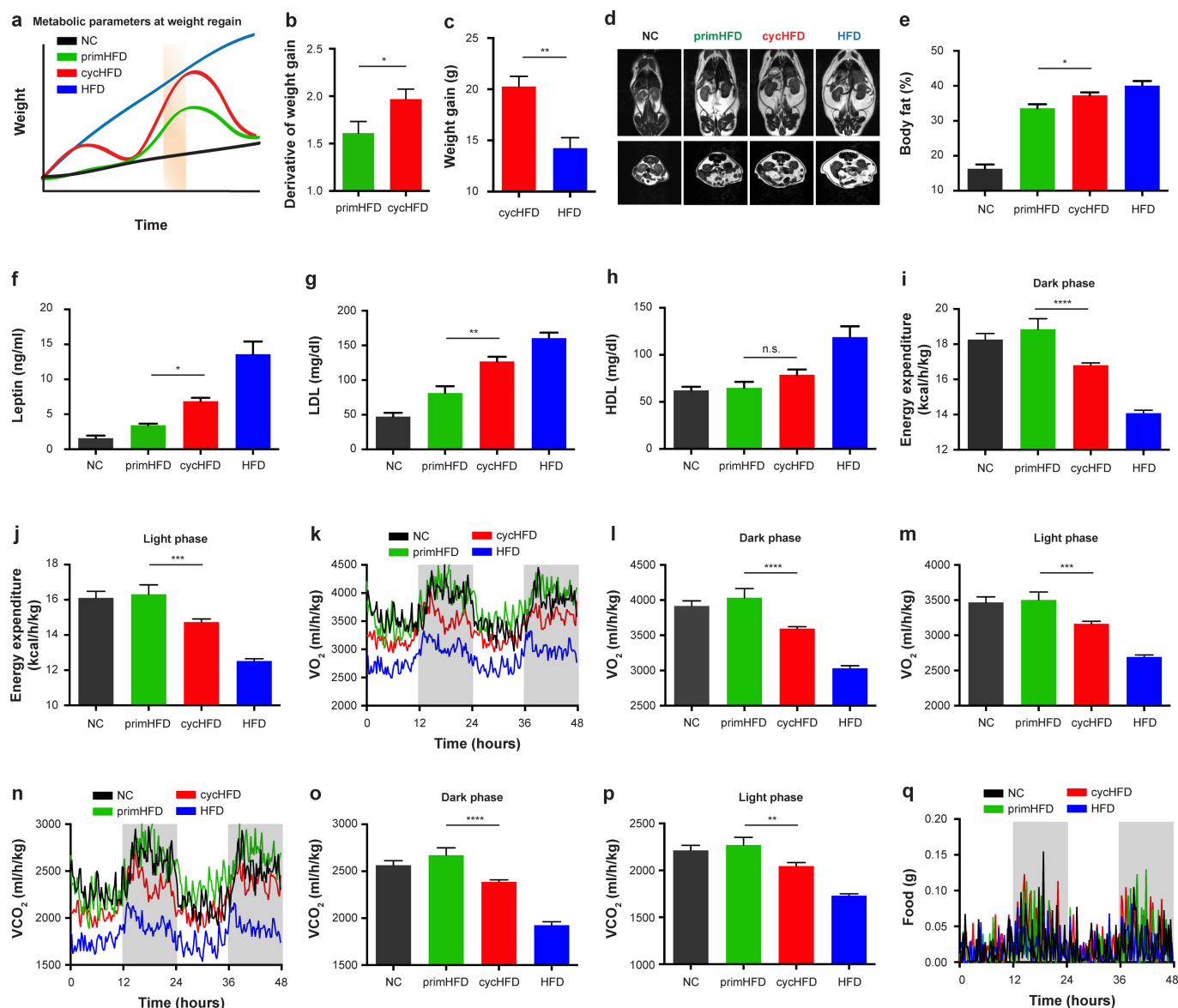
Western blot analysis. Brown adipose tissue samples were excised and washed thoroughly with PBS, homogenized in RIPA buffer containing protease inhibitors, incubated for 20 min in 4°C and centrifuged for 20 min, $14,000\text{ r.p.m.}$, at 4°C . Samples were separated on 12% acrylamide gels and transferred onto nitrocellulose membranes. Western blot analysis was performed using anti-UCP1 (M-17) polyclonal antibody (Santa Cruz, sc-6529) and donkey anti-goat antibody (Jackson ImmunoResearch, 705-035-003). Band density was calculated using ImageJ software. See Supplementary Fig. 4 for immunoblot source data.

Adipose tissue explants. Brown adipose tissue was excised and rinsed with PBS. The tissue explants were cultured with 0, 10 or $100\mu\text{M}$ of apigenin and naringenin for 24 h in DMEM medium containing 10% FBS, L-glutamine, penicillin, and streptomycin at 37°C . Explants were then collected and immediately processed for qPCR and western blot analysis as described above.

Statistical analysis. Data are expressed as mean \pm s.e.m. Comparisons between two groups were performed using two-tailed Mann-Whitney *U*-test. ANOVA was used for comparison between multiple groups. Statistical testing was performed using GraphPad Prism software. *K*-means clustering based on Pearson's correlation was used to classify the temporal behaviour of OTUs, metagenomes, and metabolites in weight-cycling mice after normalization to control mice in order to account for ageing-induced changes. *P* values <0.05 were considered significant. **P* <0.05 ; ***P* <0.01 ; ****P* <0.001 ; *****P* <0.0001 . Exact *P* values for each experiment can be found in Supplementary Table 6.

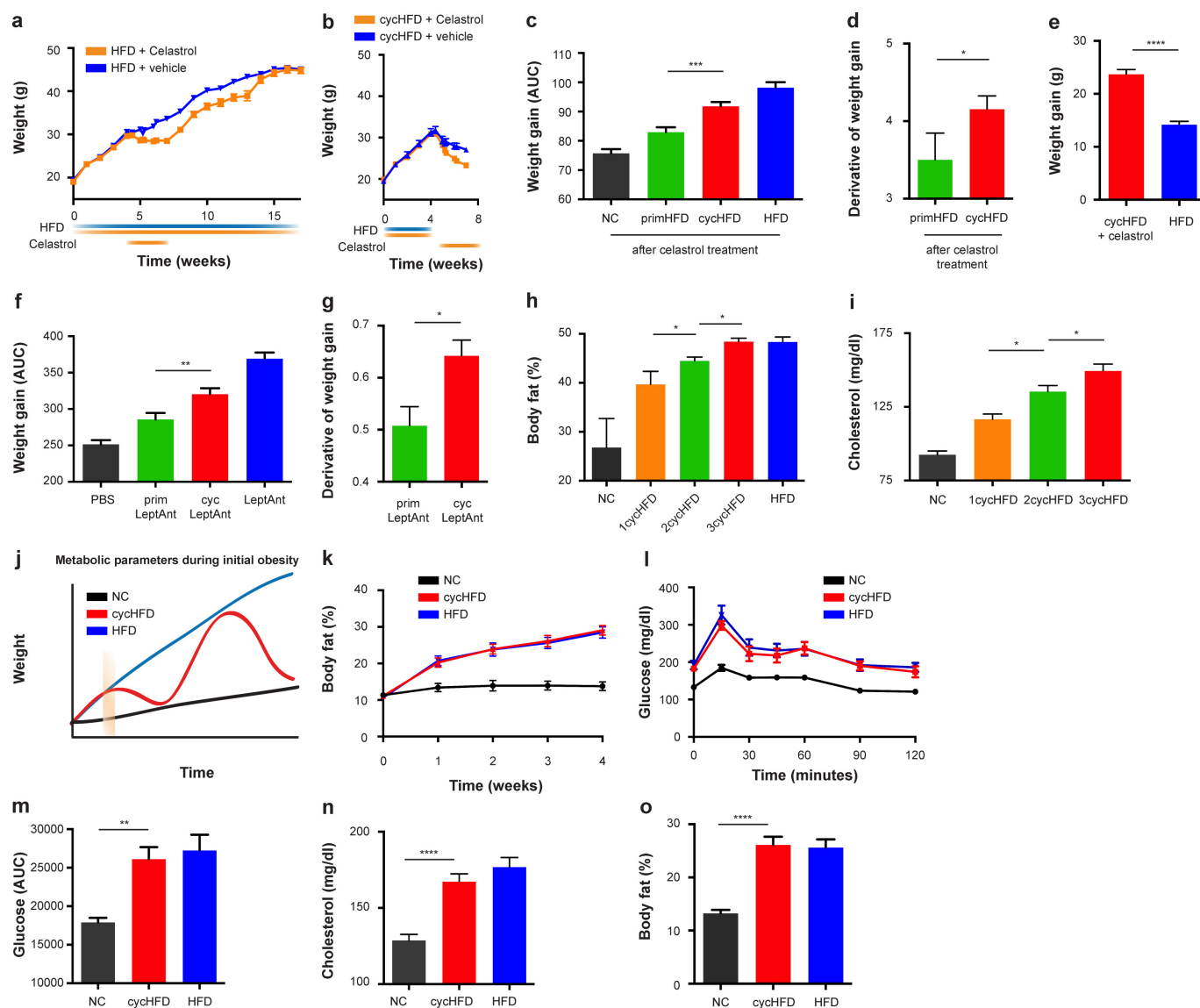
Data availability. The sequencing data has been deposited at the European Nucleotide Archive database with the accession number PRJEB17697 and are available from the corresponding authors upon request.

27. Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. & Medzhitov, R. Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* **118**, 229–241 (2004).
28. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
29. Thaïss, C. A. *et al.* Transkingdom control of microbiota diurnal oscillations promotes metabolic homeostasis. *Cell* **159**, 514–529 (2014).
30. Levy, M. *et al.* Microbiota-modulated metabolites shape the intestinal microenvironment by regulating NLRP6 inflammasome signaling. *Cell* **163**, 1428–1443 (2015).
31. Li, J. *et al.* An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).
32. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
33. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
34. Marco-Sola, S., Sammeth, M., Guigó, R. & Ribeca, P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat. Methods* **9**, 1185–1188 (2012).



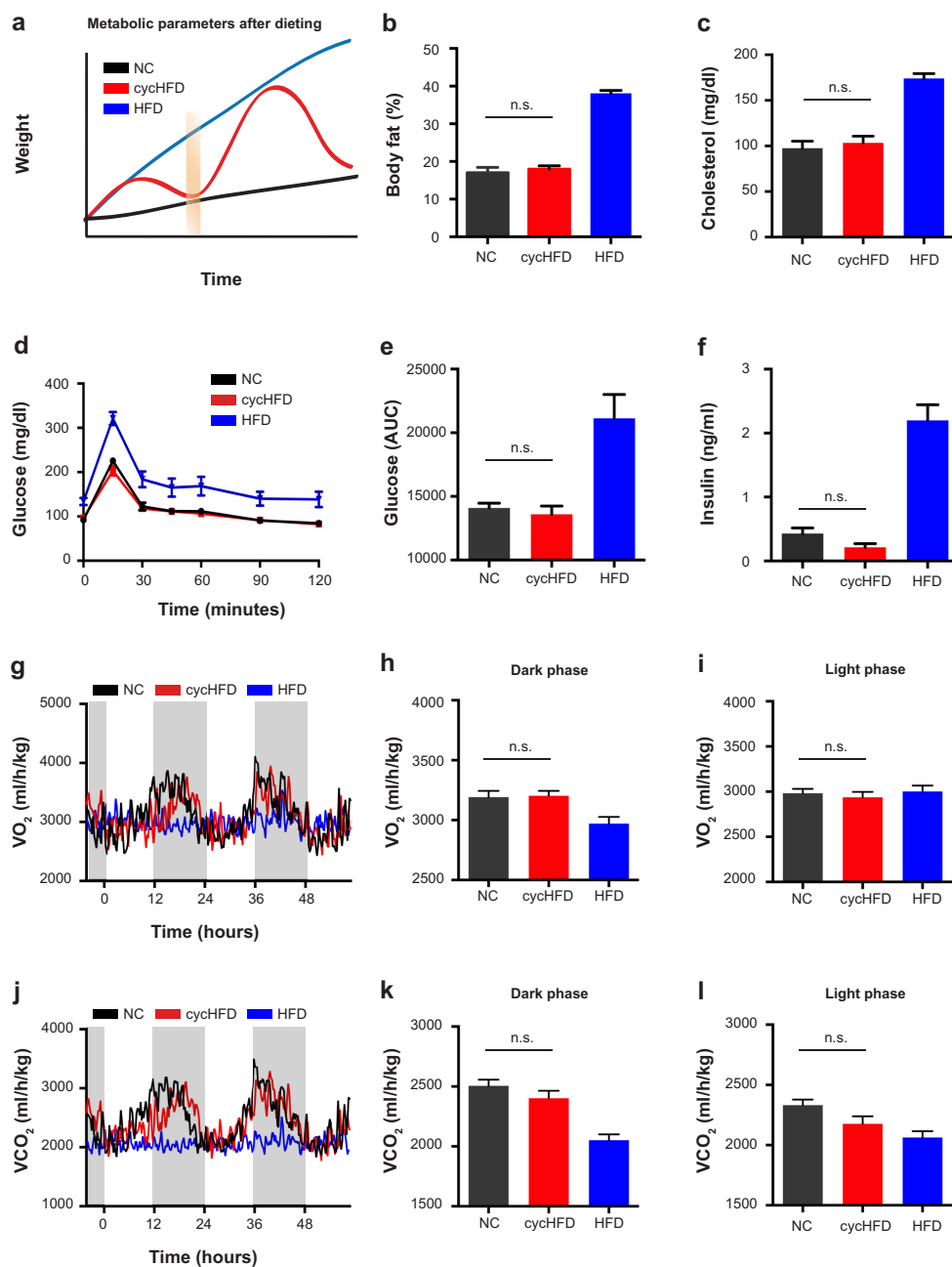
Extended Data Figure 1 | Metabolic measurements during weight regain. **a**, Schematic indicating time point of metabolic measurements. **b**, Quantification of weight regain by weight gain. **c**, Net weight gain induced by 8 weeks of HFD in weight-cycling mice and continuous HFD. **d**, **e**, Coronal (above) and axial (below) MRI scans (**d**), and quantification of body fat content (**e**). **f–h**, Serum levels of leptin (**f**), LDL (**g**), and HDL (**h**) during the second HFD exposure of mice undergoing weight cycling and controls. **i**, **j**, Quantification of dark phase (**i**) and light phase (**j**)

energy expenditure upon weight regain of weight-cycling mice. **k–q**, Representative recordings (**k**, **n**, **q**) and quantifications (**l**, **m**, **o**, **p**) of O_2 consumption (**k–m**), CO_2 consumption (**n–p**), and food intake (**q**) upon weight regain of weight-cycling mice. Experiments were repeated twice. Shown are mean \pm s.e.m. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$ by ANOVA (**e–p**) or Mann–Whitney *U*-test (**b**, **c**). See Supplementary Tables 5 and 6 for exact *n* and *P* values.



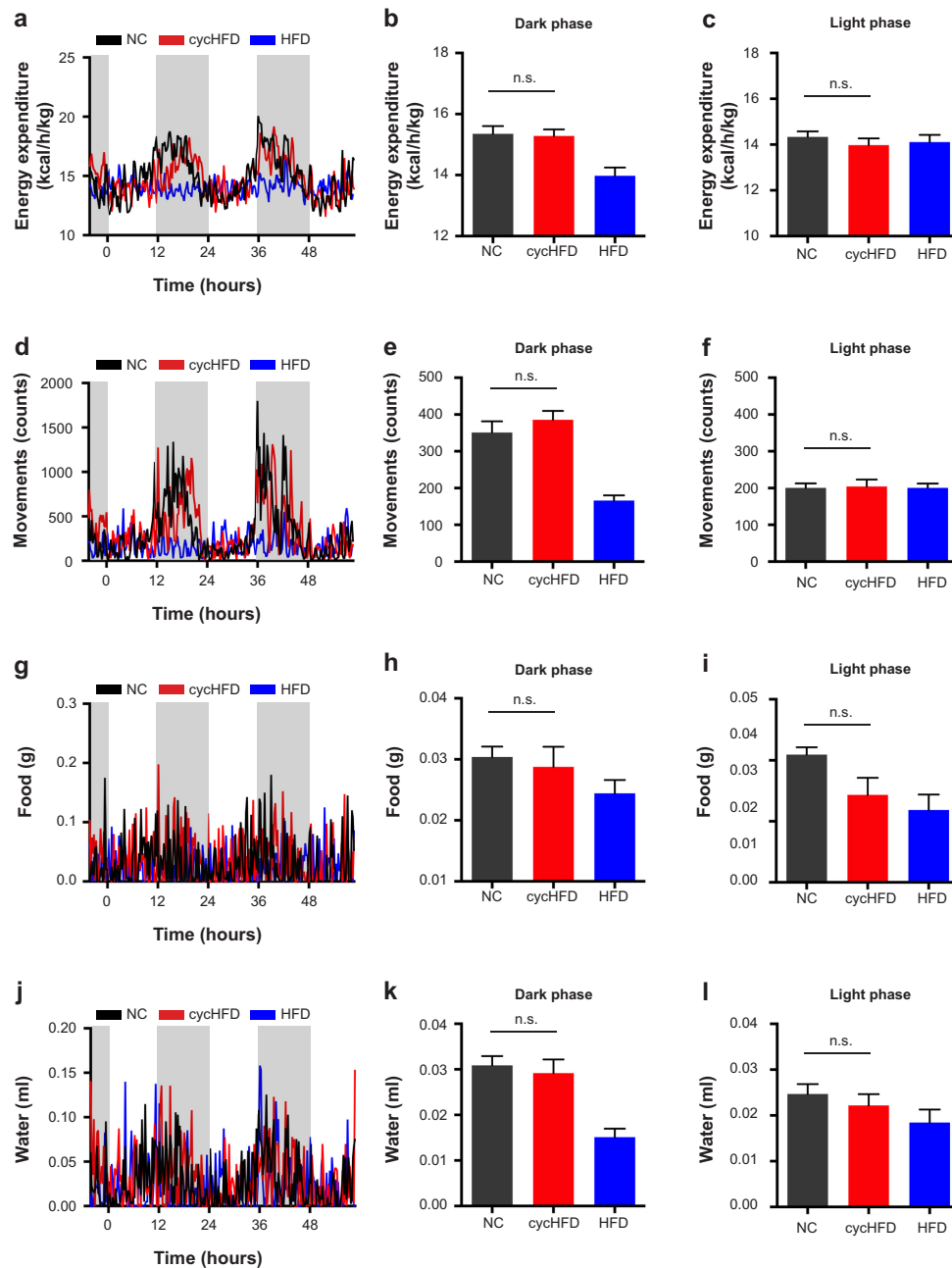
Extended Data Figure 2 | Enhanced recurrent weight gain after initial obesity. **a, b**, The effect of celastrol on weight loss in mice continuously fed a HFD (**a**) and mice with alternating diets (**b**). **c–e**, Weight regain quantification by AUC (**c**), weight regain slope (**d**), and net weight gain on HFD (**e**) by control mice and weight-cycling mice treated with celastrol to lose weight. **f, g**, Quantification of weight regain by AUC (**f**), weight regain slope (**g**) of leptin antagonist-treated weight-cycling mice and controls. **h, i**, Body fat content (**h**), and serum cholesterol levels (**i**) of weight-cycling mice undergoing a third weight cycle and controls. **j**, Schematic of the

analysed time point in **k–o**. **k**, Weight gain during 4 weeks of HFD. **l–o**, Glucose levels after glucose tolerance test (**l**), glucose level quantification (**m**), serum cholesterol levels (**n**), and quantification of body fat content (**o**) in weight-cycling mice during initial obesity and controls. Experiments were repeated twice. Shown are mean \pm s.e.m. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$ by ANOVA (**c, f, h, i, m, n, o**) or Mann–Whitney U -test (**d, e, g**). See Supplementary Tables 5 and 6 for exact n and P values.



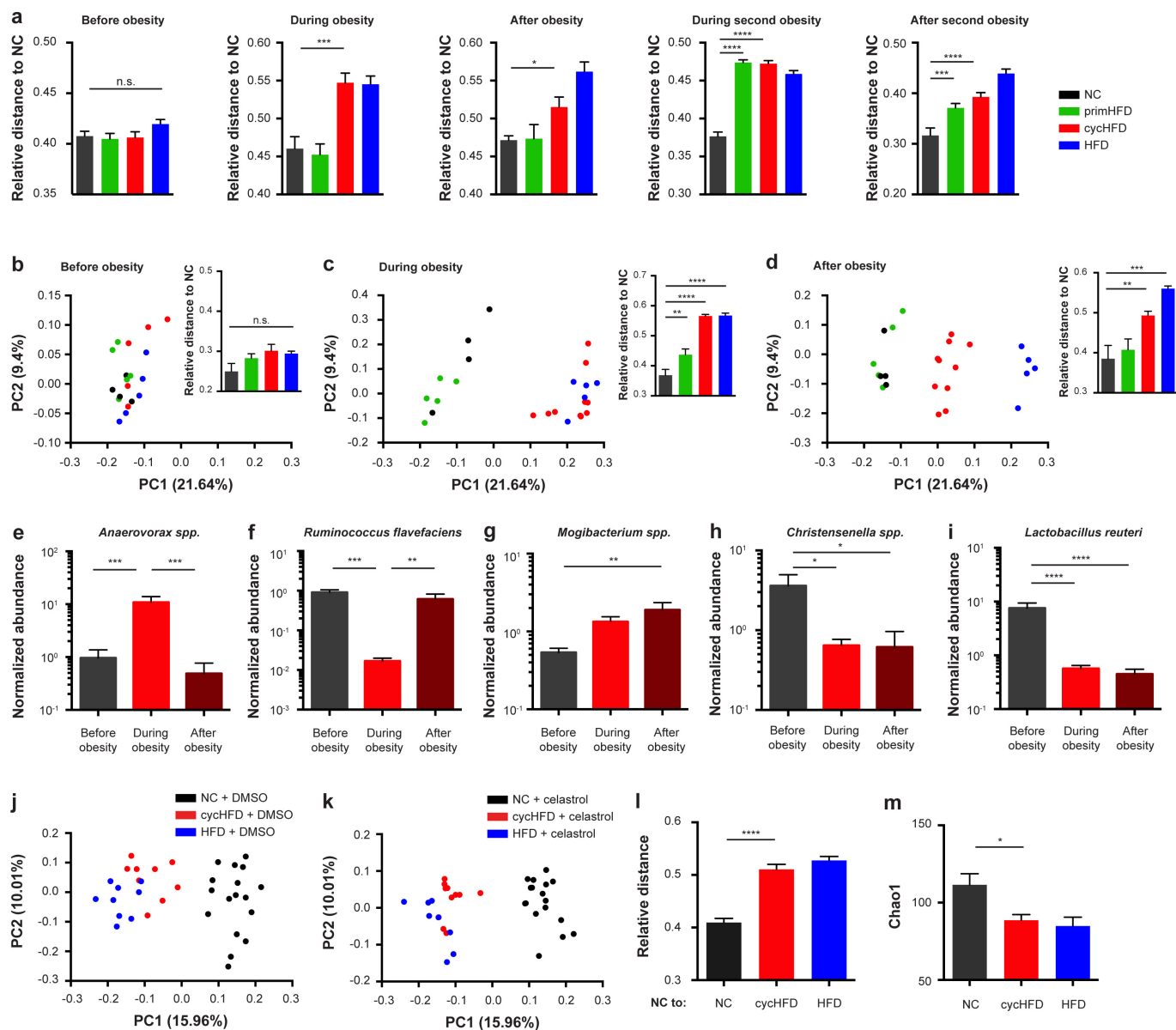
Extended Data Figure 3 | Recovery of metabolic parameters after dieting. **a**, Schematic of the analysed nadir time point. **b–f**, Body fat content (**b**), serum cholesterol levels (**c**), glucose levels after glucose tolerance test (**d**), glucose level quantification (**e**), and serum insulin levels (**f**) in weight-cycling mice upon return to normal weight.

g–l, Representative recordings (**g**, **j**) and quantifications (**h**, **i**, **k**, **l**) of O_2 and CO_2 consumption by weight-cycling mice upon return to normal weight and controls. Experiments were repeated twice. Data are mean \pm s.e.m. n.s., not significant by ANOVA. See Supplementary Table 5 for exact n values.



Extended Data Figure 4 | Metabolic measurements after dieting. Representative recordings (a, d, g, j), dark phase (b, e, h, k), and light phase (c, f, i, l) quantifications of energy expenditure (a–c), physical activity (d–f), food intake (g–i), and water consumption (j–l) of

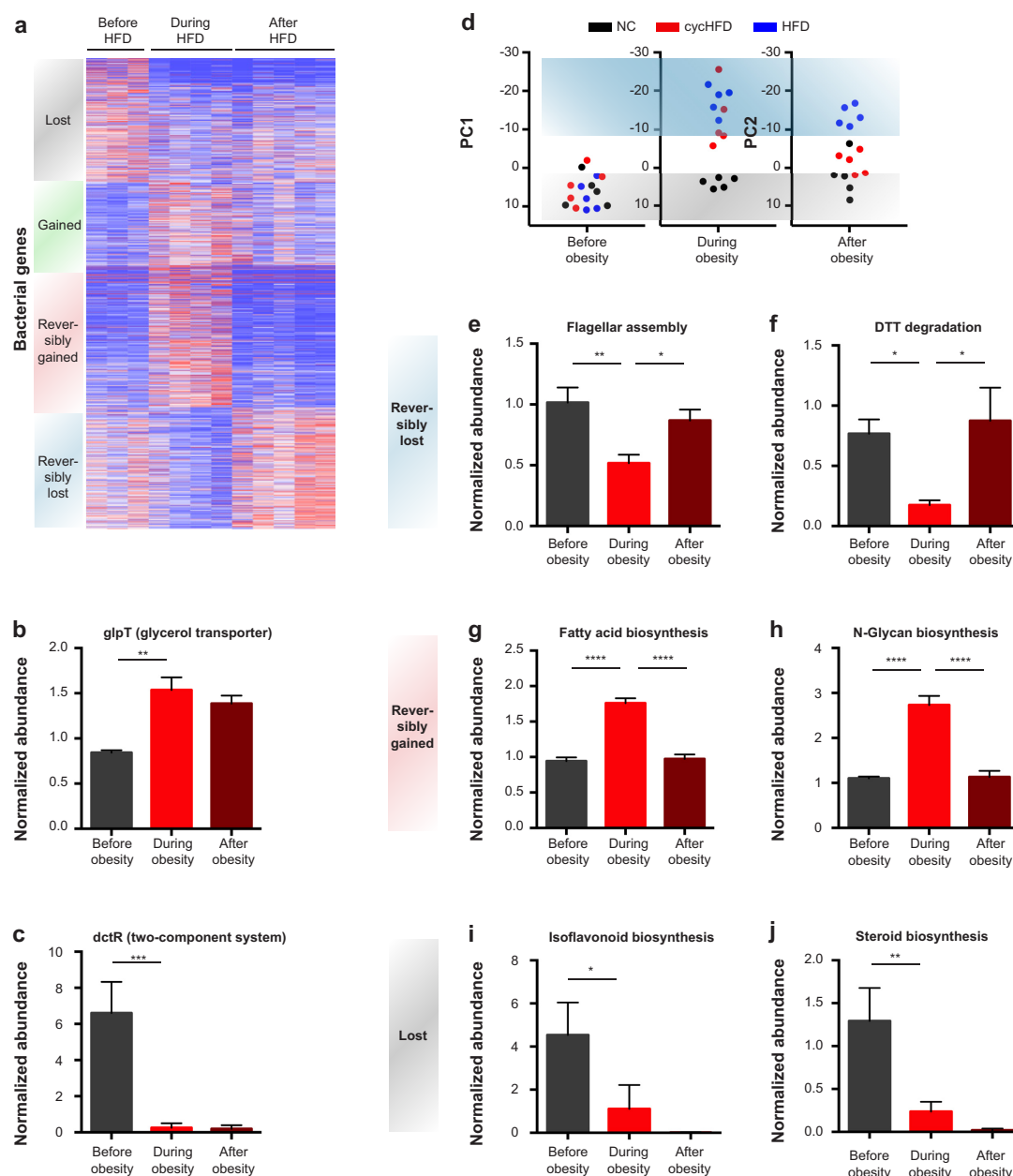
weight-cycling mice during weight regain and controls. Experiments were repeated twice. Shown are mean \pm s.e.m. n.s., not significant by ANOVA. See Supplementary Table 5 for exact *n* values.



Extended Data Figure 5 | Persistent microbiome changes after dieting.

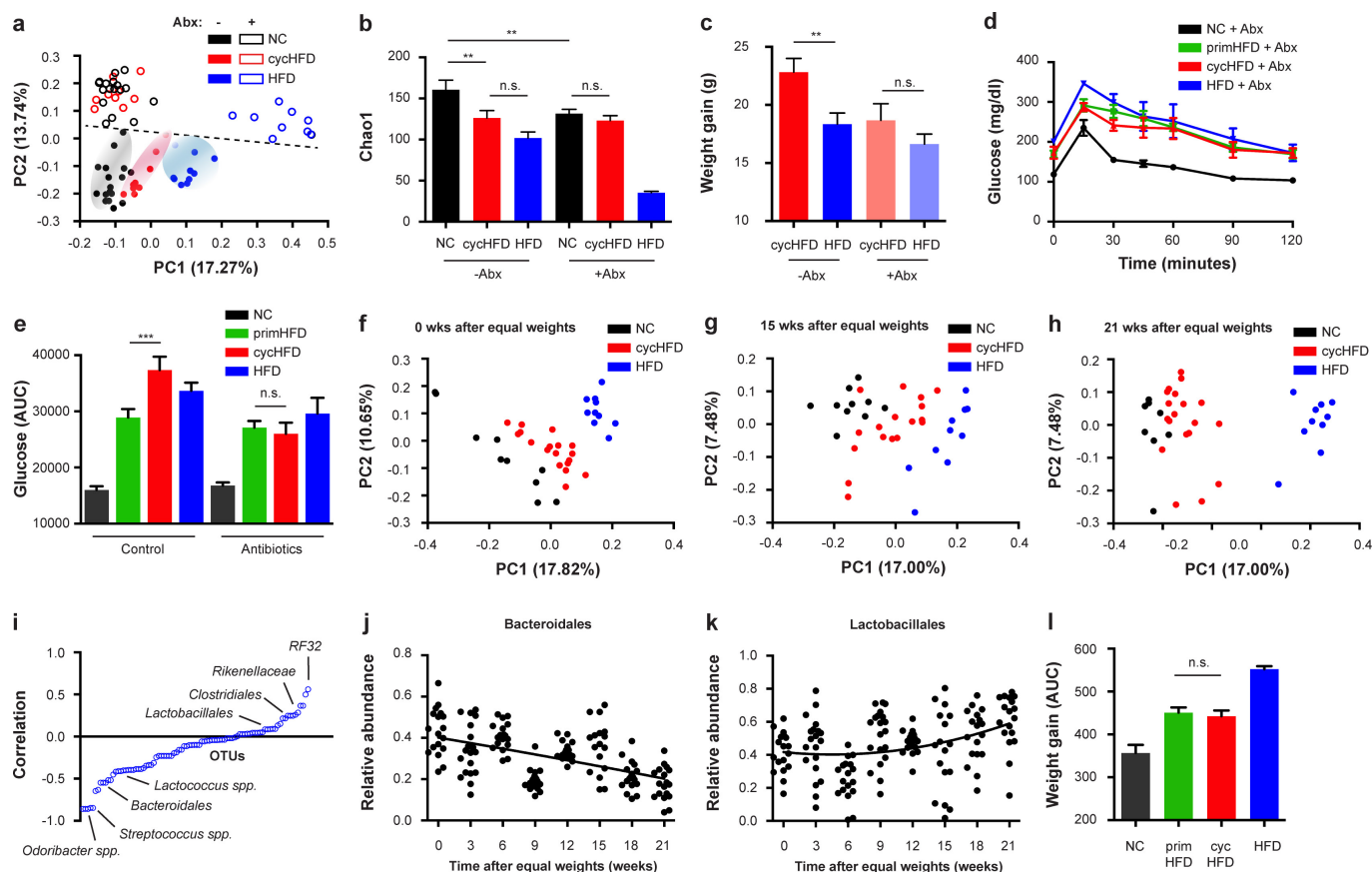
a, Quantification of UniFrac distances of weight-cycling mice from NC controls at the indicated time points. Plots correspond to the PCoA analyses in Fig. 2b–f. **b–d**, PCoA analyses and distance quantification (insets) of V3/V4-targeted 16S sequencing of mice before (**b**), during (**c**), and after (**d**) diet-induced obesity and subsequent weight loss. **e–i**, Examples of OTUs whose abundance does (**e**, **f**) or does not (**g–i**)

recover after dieting. **j–m**, PCoA analyses (**j**, **k**), UniFrac distance (**l**) and alpha diversity (**m**) at the nadir time point of post-dieting mice that had received celastrol to accelerate weight loss. Experiments were repeated twice. Data are mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$ by ANOVA. See Supplementary Tables 5 and 6 for exact n and P values.



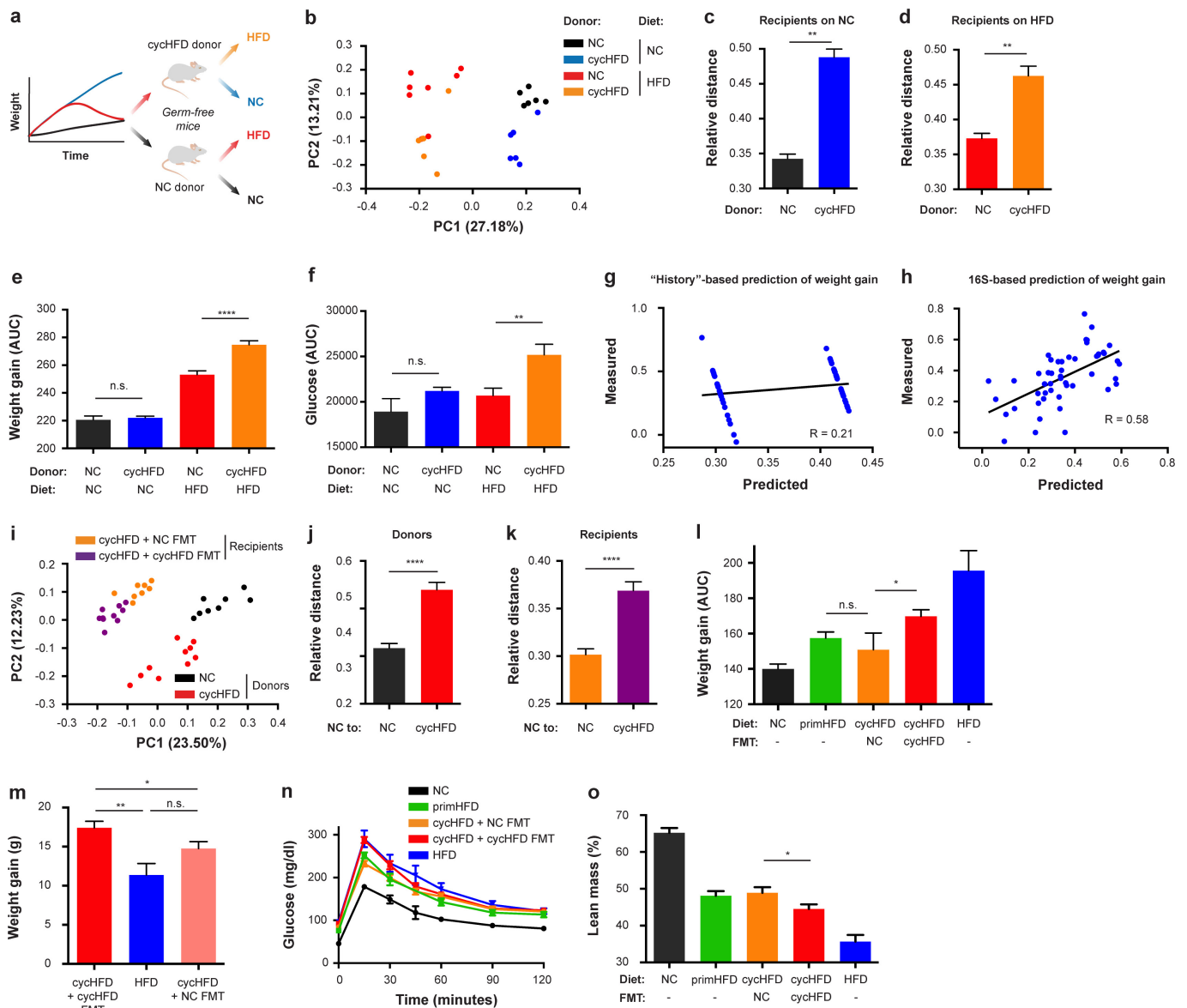
Extended Data Figure 6 | Persistent metagenomic changes after dieting. **a**, Heatmap of normalized gene abundance in the microbiota of mice before, during, and after obesity. **b**, **c**, Examples of genes whose abundance does not recover after dieting. **d**, PCA of bacterial KEGG modules over time in weight-cycling mice and controls. **e–j**, Examples of

KEGG pathways whose abundance is reversibly decreased (**e**, **f**), reversibly increased (**g**, **h**), or persistently decreased (**i**, **j**) during obesity and dieting. Data are from one experiment. Data are mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$ by ANOVA. See Supplementary Tables 5 and 6 for exact n and P values.



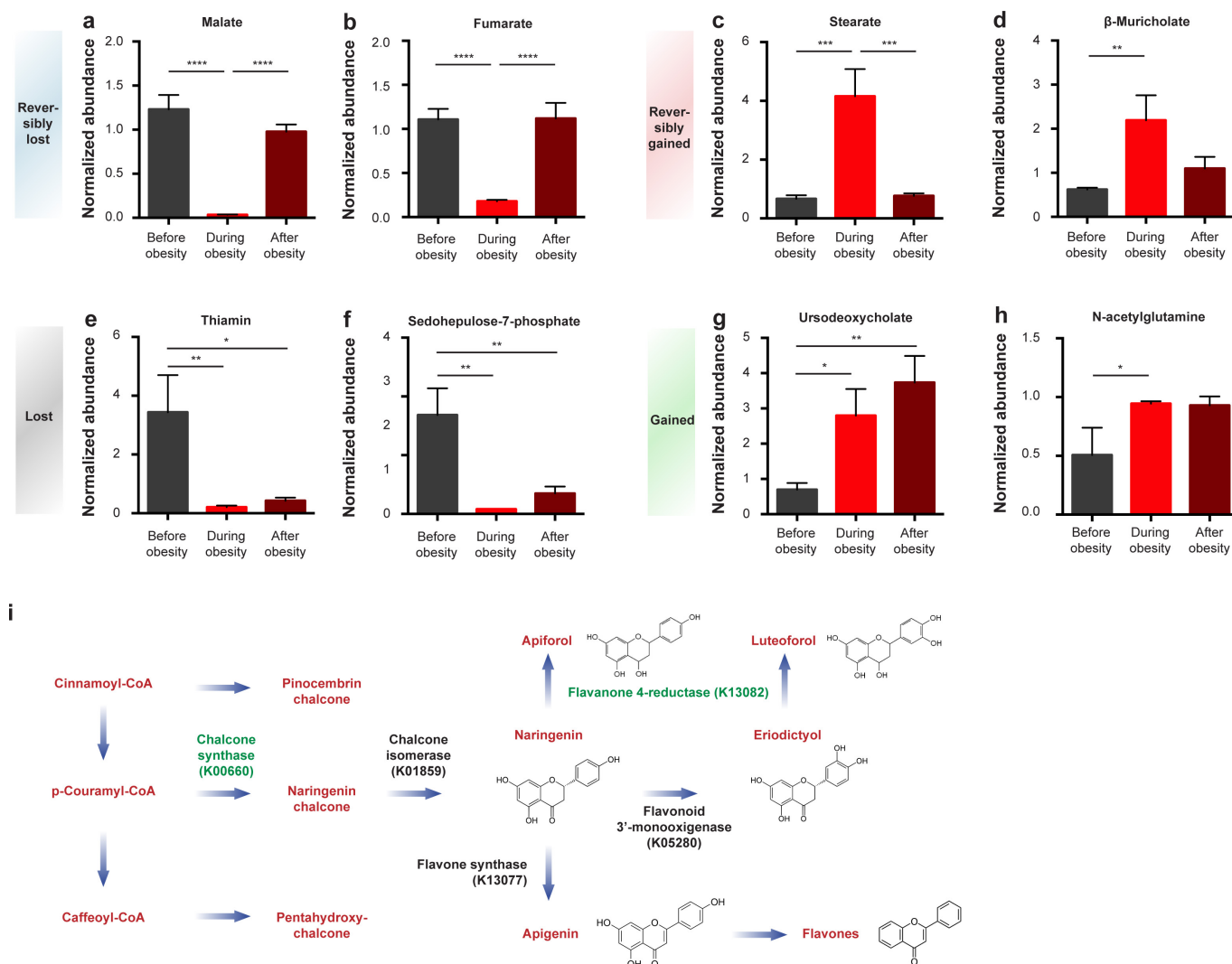
Extended Data Figure 7 | The post-dieting microbiota drives enhanced recurrent obesity. **a, b**, PCoA (**a**) and alpha diversity (**b**) of faecal microbiota after dieting (nadir time point at week 8) from mice with or without antibiotic treatment during weight loss. **c**, Net weight gain induced by 8 weeks of HFD in weight-cycling mice or continuous HFD control with or without antibiotic treatment between weeks 4 and 8. **d, e**, Glucose levels after oral glucose tolerance test (**d**) and glucose level quantification after glucose tolerance test (**e**) during weight regain in mice with or without antibiotic treatment during weight loss ($n = 4-10$).

f-h, PCoA of faecal microbiota from formerly obese mice and controls at the time of dieting-induced weight normalization (**f**), 15 weeks after weight normalization (**g**), and 21 weeks after weight normalization (**h**). **i-k**, Correlation analysis (**i**) and examples (**j, k**) of microbial taxa undergoing gradual normalization in abundance over a period of 21 weeks after weight normalization. **l**, Quantification of secondary weight gain after microbiota normalization. Experiments were repeated twice. Data are mean \pm s.e.m. ** $P < 0.01$, *** $P < 0.001$ by ANOVA. See Supplementary Tables 5 and 6 for exact n and P values.



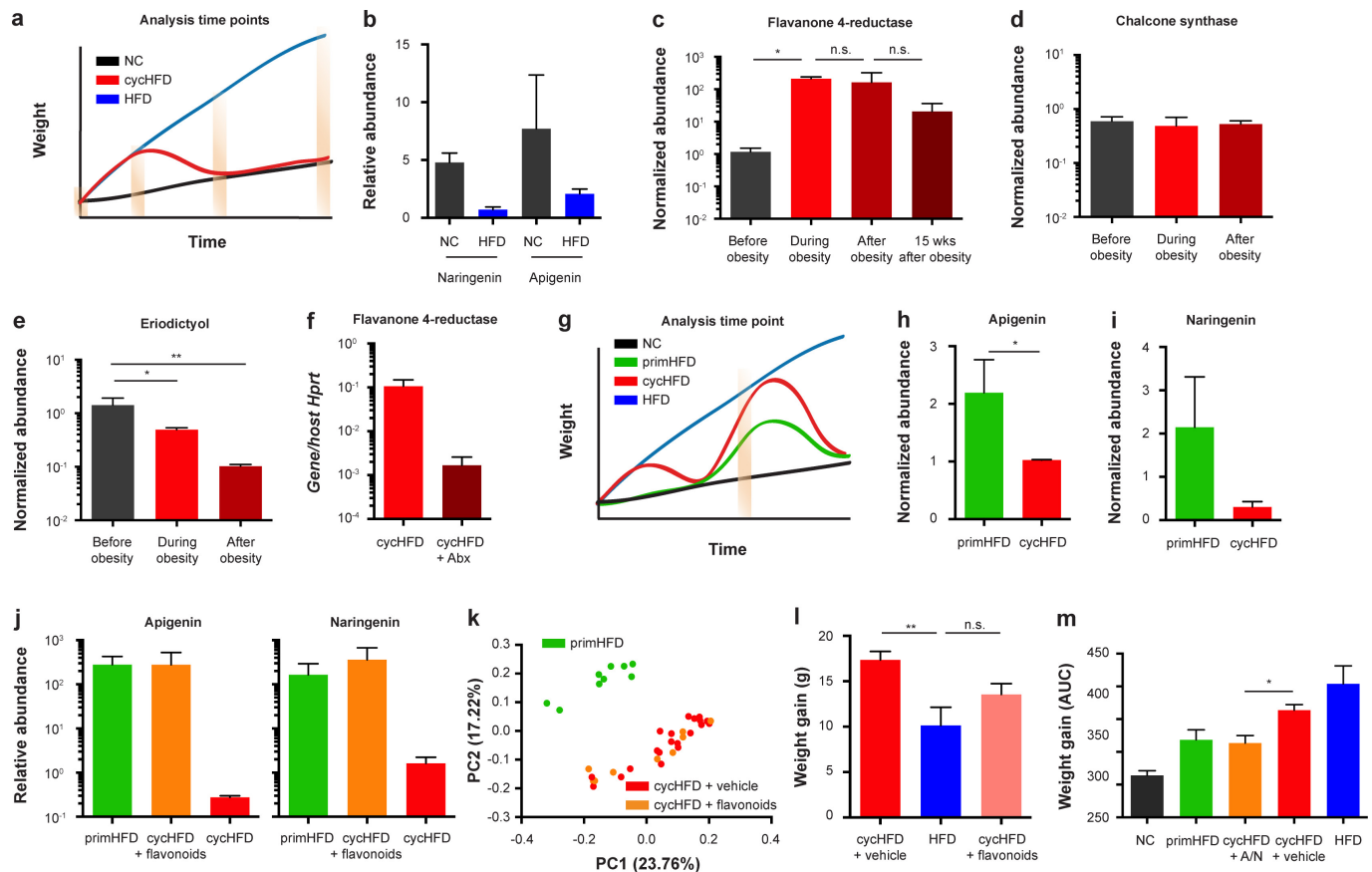
Extended Data Figure 8 | Transfer, prediction and treatment of weight regain by microbiome features. a, Schematic of microbiota transfer to germ-free mice after dieting. Recipients were fed either a HFD or NC. b–d, PCoA of recipient microbiota (b) and relative UniFrac distances to NC controls (c, d) of germ-free mice one week after transplantation with microbiota from weight-cycling mice or controls, and fed either NC (c) or a HFD (d). e, f, Quantifications of weight gain (e) and blood glucose after glucose tolerance test (f) in germ-free recipients of microbiota from weight cycling mice or controls. g, h, Correlation of predicted and measured weight gain when prediction is based solely on inferred history of

obesity (g) or solely on 16S sequencing (h). i–k, PCoA of faecal microbiota (i) and relative UniFrac distances between donors (j) and recipients (k) two weeks after the onset of daily FMT from cycHFD or NC mice to mice undergoing weight cycling. l–o, Quantification of secondary weight gain (l), net weight gain induced by 8 weeks of HFD feeding (m), glucose levels after glucose tolerance test (n) and lean mass (o) in weight-cycling mice and controls with or without FMT. Experiments were repeated twice. Data are mean \pm s.e.m. n.s., not significant. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$ by ANOVA (e, f, i, m, o) or Mann-Whitney U -test (c, d, j, k). See Supplementary Tables 5 and 6 for exact n and P values.



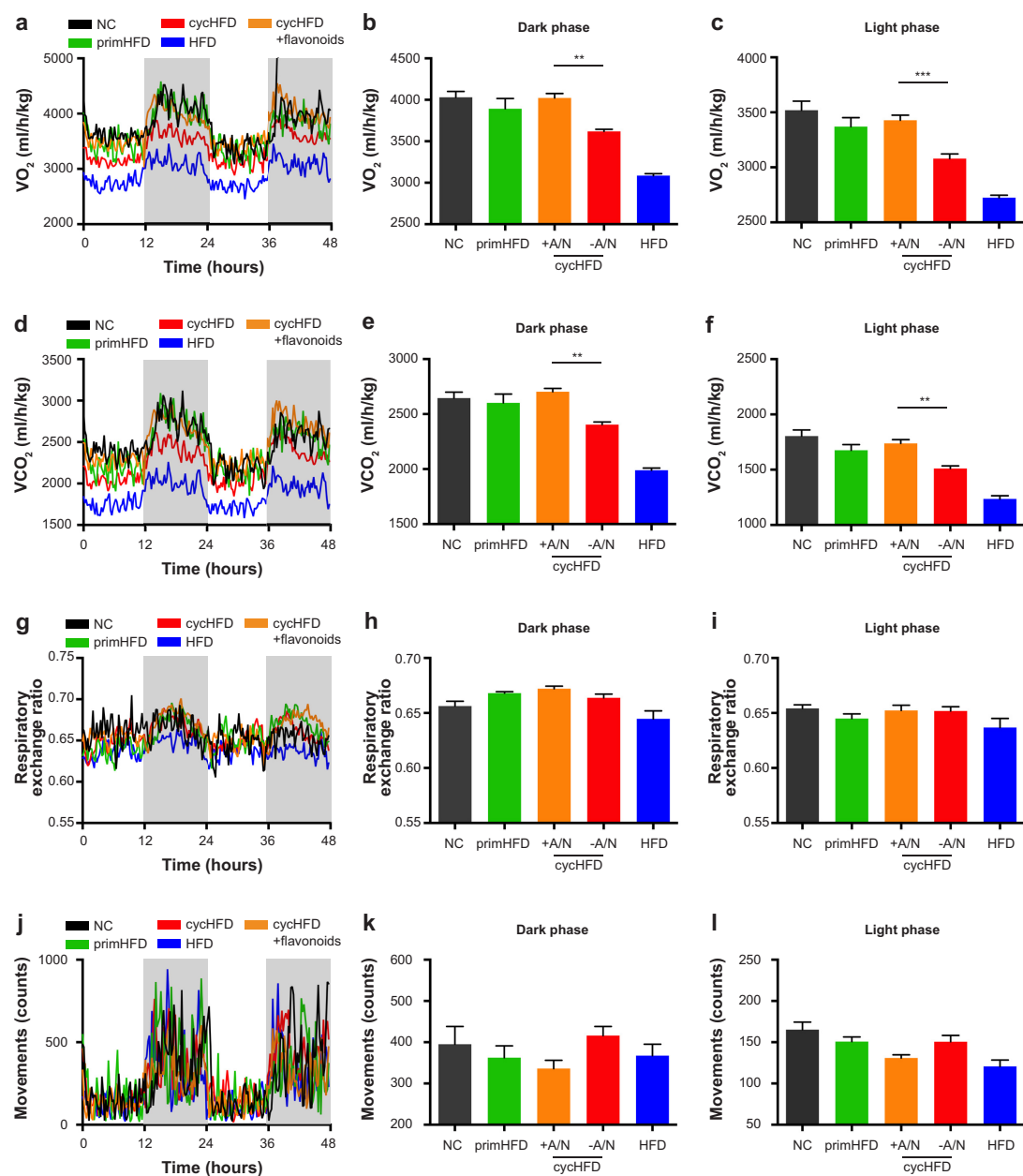
Extended Data Figure 9 | Persistent metabolomic changes after dieting. **a–h**, Examples of metabolites whose abundance is reversibly decreased (**a**, **b**), reversibly increased (**c**, **d**), persistently decreased (**e**, **f**), or persistently increased (**g**, **h**) during obesity and dieting. **i**, Schematic of flavonoid biosynthetic pathways leading to the production and conversion

of naringenin. KEGG IDs of key enzymes are indicated. Genes found in our metagenomic dataset are indicated in green. Data are from one experiment. Shown are mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$ by ANOVA. See Supplementary Tables 5 and 6 for exact n and P values.



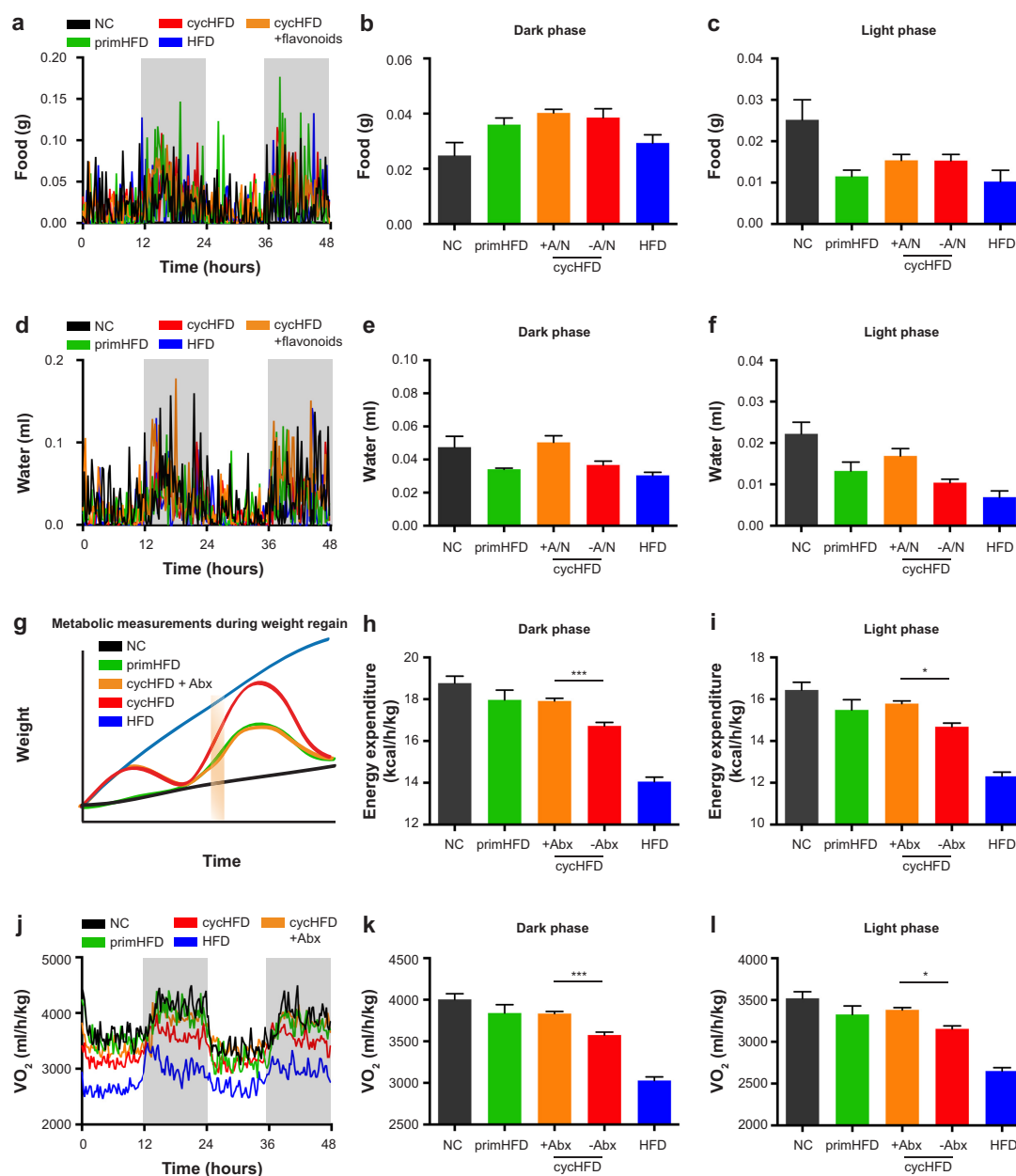
Extended Data Figure 10 | Microbiota control of post-dieting metabolic complications through intestinal flavonoids. **a**, Schematic showing sampling times in obesity/recovery cycle experiment. **b**, Dietary flavonoids in NC and HFD. **c–e**, Abundance of flavanone 4-reductase (**c**), chalcone synthase (**d**), and eriodictyol (**e**) over time in the faeces of mice undergoing weight cycling normalized to controls. **f**, Quantification of flavanone 4-reductase levels in faecal DNA relative to host DNA in weight-cycling mice at the end of the weight loss period, with or without antibiotic treatment during weight loss. **g**, Schematic of sampling time upon weight

regain. **h, i**, Abundance of apigenin (**h**) and naringenin (**i**) in the faeces of mice undergoing post-dieting weight regain and controls. **j–m**, Flavonoid levels (**j**), PCoA of faecal microbiota (**k**), net weight gain induced by 8 weeks of HFD (**l**), and weight regain quantification by AUC (**m**) of weight-cycling mice supplemented with apigenin and naringenin during the weight regain. Data are from one (**a–k**) or two (**l, m**) experiments. Shown are mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$ by ANOVA (**c, e, l**) or Mann–Whitney U -test (**h, m**). See Supplementary Tables 5 and 6 for exact n and P values.



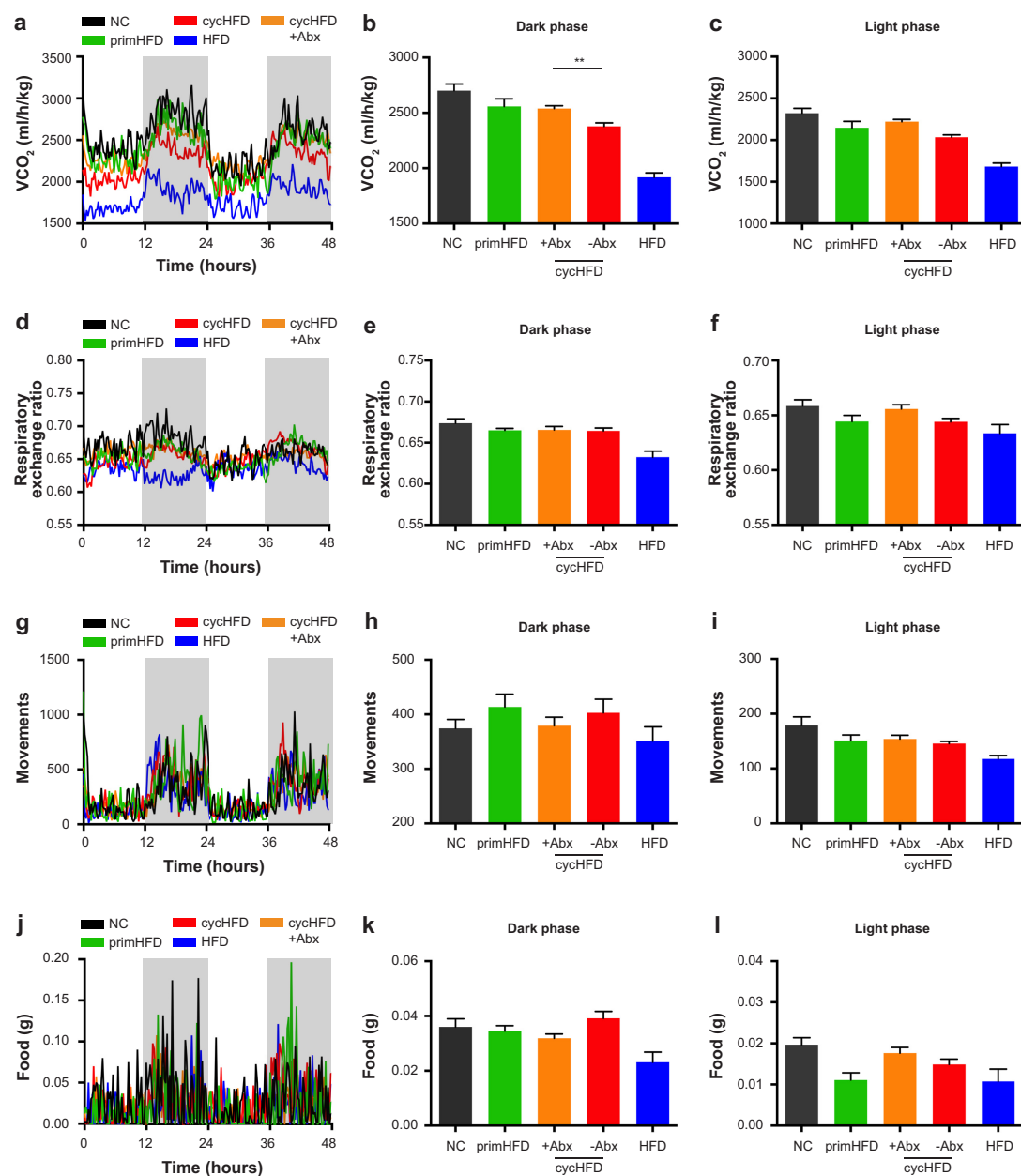
Extended Data Figure 11 | Metabolic measurements in flavonoid-treated mice. Representative recordings (a, d, g, j), dark phase quantifications (b, e, h, k), and light phase quantifications (c, f, i, l) of O_2 consumption (a–c), CO_2 consumption (d–f), respiratory exchange ratio (g–i), and physical activity (j–l) of weight-cycling mice with or without

supplementation of apigenin and naringenin (A/N) during weight regain. Data are from one experiment. Data are mean \pm s.e.m. ** $P < 0.01$, *** $P < 0.001$ by ANOVA. See Supplementary Tables 5 and 6 for exact n and P values.



Extended Data Figure 12 | Metabolic measurements in flavonoid- and antibiotic-treated mice. Representative recordings (**a**, **d**), dark phase quantifications (**b**, **e**), and light phase quantifications (**c**, **f**) of food (**a–c**) and water consumption (**d–f**), of weight-cycling mice with or without supplementation of apigenin and naringenin (A/N) during weight regain. **g**, Schematic indicating time of metabolic measurements during the weight regain phase. **h**, **i**, Quantifications of energy expenditure in

weight-cycling mice with or without antibiotic treatment during weight loss. **j–l**, Representative recording (**j**) and quantifications (**k**, **l**) of O₂ consumption by weight-cycling mice with or without antibiotic treatment (Abx) during weight loss. Data are from one experiment. Shown are mean \pm s.e.m. * P < 0.05, *** P < 0.001 by ANOVA. See Supplementary Tables 5 and 6 for exact n and P values.



Extended Data Figure 13 | Metabolic measurements in antibiotic-treated mice. Representative recordings (a, d, g, f), dark phase quantifications (b, e, h, k), and light phase quantifications (c, f, i, l) of CO_2 consumption (a–c), respiratory exchange ratio (d–f), physical activity

(g–i) and food intake (j–l) by weight-cycling mice with or without antibiotic treatment (Abx) during weight loss. Data are from one experiment. Data are mean \pm s.e.m. *** $P < 0.01$ by ANOVA. See Supplementary Tables 5 and 6 for exact n and P values.

Early dissemination seeds metastasis in breast cancer

Hedayatollah Hosseini¹, Milan M. S. Obradović^{1†}, Martin Hoffmann², Kathryn L. Harper³, Maria Soledad Sosa^{3†}, Melanie Werner-Klein⁴, Lahiri Kanth Nanduri^{1†}, Christian Werno², Carolin Ehrlich¹, Matthias Maneck¹, Nina Patwary¹, Gundula Haunschild¹, Miodrag Gužvić¹, Christian Reimelt¹, Michael Grauvogel⁵, Norbert Eichner⁶, Florian Weber⁷, Andreas D. Hartkopf⁸, Florin-Andrei Taran⁸, Sara Y. Brucker⁸, Tanja Fehm⁹, Brigitte Rack¹⁰, Stefan Buchholz¹¹, Rainer Spang⁵, Gunter Meister⁶, Julio A. Aguirre-Ghiso³ & Christoph A. Klein^{1,2}

Accumulating data suggest that metastatic dissemination often occurs early during tumour formation, but the mechanisms of early metastatic spread have not yet been addressed. Here, by studying metastasis in a HER2-driven mouse breast cancer model, we show that progesterone-induced signalling triggers migration of cancer cells from early lesions shortly after HER2 activation, but promotes proliferation in advanced primary tumour cells. The switch from migration to proliferation was regulated by increased HER2 expression and tumour-cell density involving microRNA-mediated progesterone receptor downregulation, and was reversible. Cells from early, low-density lesions displayed more stemness features, migrated more and founded more metastases than cells from dense, advanced tumours. Notably, we found that at least 80% of metastases were derived from early disseminated cancer cells. Karyotypic and phenotypic analysis of human disseminated cancer cells and primary tumours corroborated the relevance of these findings for human metastatic dissemination.

Systemic cancer (the dissemination and subsequent distant outgrowth of cells from a solid tumour) occurs in two phases: a clinically latent stage of hidden cancer spread, and then manifest metastasis. Manifest metastasis remains mostly incurable. The period of clinically undetectable minimal residual disease, defined by disseminated cancer cells (DCCs) that are left behind after primary tumour surgery, offers a time window to prevent metastasis^{1,2}. However, only circumstantial knowledge is available about minimal residual disease, and consequently, systemic (adjuvant) therapies improve outcome in only about 20% of patients^{3,4}. This situation indicates that our current understanding of early systemic cancer is insufficient to prevent metastasis.

The first direct evidence for a characteristic biology of early disseminated cancer and minimal residual disease came from analyses of DCCs isolated from bone marrow of patients with breast cancer before (M0 stage, according to Union for International Cancer Control guidelines) and after (M1 stage) manifestation of metastasis^{5,6}, indicating that M0-DCCs might have disseminated early and evolved in parallel with the primary tumour⁷. Studies in transgenic mouse models^{8–10} and in patients with pre-malignant lesions or *in situ* carcinomas^{8,11,12} corroborated this concept but the relevance of DCCs remains contested¹³.

We therefore addressed the issue of breast-cancer-cell dissemination soon after cancer initiation and investigated whether mechanisms exist that reduce metastatic seeding from advanced cancers. Finally, we investigated whether early DCCs (eDCCs) are able to form metastases. We report on a mechanism involving cell density, HER2 and PGR signalling that reconciles early and late dissemination models.

PGR and HER2 regulate gene expression in early lesions

In BALB-NeuT mice, dissemination starts shortly after expression of the *Her2* transgene (also known as *ErbB2*) at puberty (around 4 weeks of age), when the first hyperplastic lesions become apparent⁸. From 4–9 weeks of age, we observed micro-invasion⁸, and a sharp decline in the ratio of DCCs to total tumour area (a measure of cell numbers at risk of dissemination) during primary tumour growth (Extended Data Fig. 1a). The genetic program governing dissemination from early lesions in microdissected tissue samples (Extended Data Fig. 1b and Supplementary Table 1) showed a signature gene expression profile compared to healthy mammary glands, primary tumours and lung metastases (Fig. 1a). We defined 1,278 gene transcripts unique to early lesions of which 300 were highly conserved between mouse and human (Supplementary Data 1).

We confirmed differential expression of selected transcripts by quantitative PCR (qPCR) (Extended Data Fig. 1c) and analysed transcript expression of steroid hormone receptors (strong candidate regulators; Supplementary Tables 2, 3), all of which, except *Esr1* (also known as *ERalpha*), showed the highest expression in early lesions (Extended Data Fig. 1d). When we assessed the expression of *Ahnak*, *Baz2a*, *Nfatc3*, *Nr3c1* and *Nr3c2* genes, which were used as surrogate markers of the early lesion signature (Fig. 1b, Supplementary Tables 2, 3 and Supplementary Data 1), progesterone was the only steroid hormone that activated a similar expression profile (Extended Data Fig. 1e), but only in early lesion cells of 9-week-old BALB-NeuT mice, and not in wild-type mammary or primary tumour cells (Fig. 1c). Moreover, the expression of progesterone receptor B (PGR-B), which is the main isoform expressed for mammary gland development^{14–16}, correlated

¹Experimental Medicine and Therapy Research, University of Regensburg, 93053 Regensburg, Germany. ²Project group 'Personalized Tumour Therapy', Fraunhofer Institute for Toxicology und Experimental Medicine, 93053 Regensburg, Germany. ³Division of Hematology and Oncology, Department of Medicine, Department of Otolaryngology, Department of Oncological Sciences, Tisch Cancer Institute, Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York 10029, USA. ⁴Institute of Immunology, University of Regensburg, 93053 Regensburg, Germany. ⁵Department of Statistical Bioinformatics, Institute of Functional Genomics, University of Regensburg, 93053 Regensburg, Germany. ⁶Biochemistry Center Regensburg (BZR), Laboratory for RNA Biology, University of Regensburg, 93053 Regensburg, Germany. ⁷Institute of Pathology, University of Regensburg, 93053 Regensburg, Germany. ⁸Department of Gynecology and Obstetrics, University of Tübingen, 72076 Tübingen, Germany. ⁹Department of Gynecology and Obstetrics, University of Düsseldorf, 40225 Düsseldorf, Germany. ¹⁰Department of Gynecology and Obstetrics, University Munich, 80337 Munich, Germany. ¹¹Department of Gynecology and Obstetrics, University Medical Center Regensburg, 93053 Regensburg, Germany. ¹²Present addresses: Tumor Heterogeneity, Metastasis and Resistance, Department of Biomedicine, University of Basel, University Hospital Basel, CH-4031 Basel, Switzerland (M.M.S.O.); Department of Pharmacological Sciences, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York 10029, USA (M.S.S.); Department of Gastrointestinal, Thoracic and Vascular Surgery, Medizinische Fakultät Carl Gustav Carus, Technische Universität Dresden, 01307 Dresden, Germany (L.K.N.).

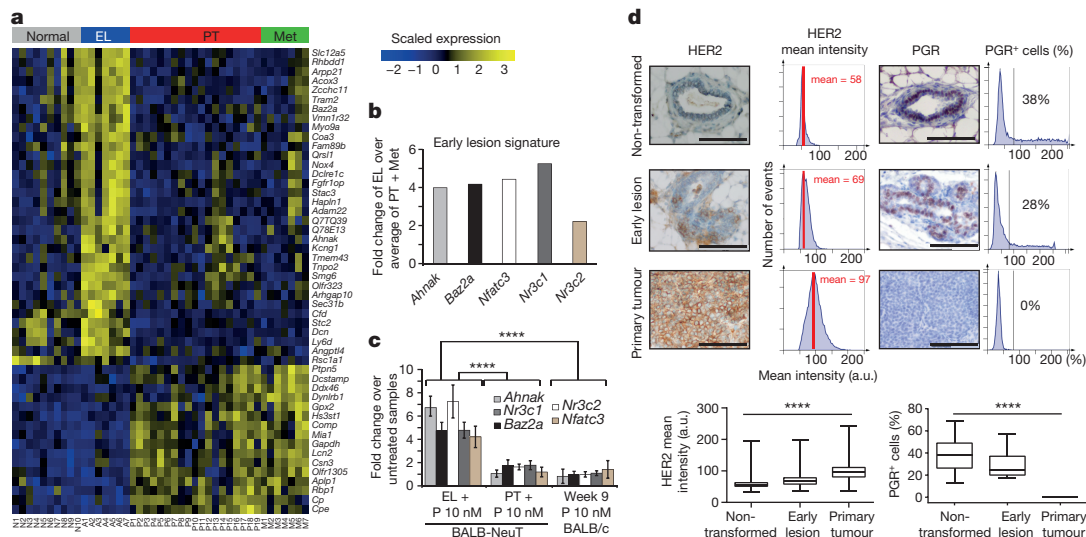


Figure 1 | Identification of a gene expression signature linked to early dissemination. **a**, Heatmap of genes that were differentially expressed in different sample types: normal mammary glands from BALB/c (N1–N10), early lesions (EL, A1–A7), primary tumours (PT, P1–P19) and lung metastases (Met, M1–M7) from BALB–NeuT mice. Yellow, upregulation; blue, downregulation. **b**, Five-gene surrogate signature (qPCR) for the early lesion profile. **c**, Progesterone (P) activates the early lesion signature *in vitro*. Data are mean \pm s.d. **d**, TissueFAX

with low to moderate HER2 expression in early lesions and wild-type mammary tissue (Extended Data Fig. 1f). We therefore quantified HER2 and PGR protein expression in mammary cells displaying normal, hyperplastic or advanced tumour morphology. The number and staining intensity of HER2⁺ cells steadily increased with advancing tumour morphology, in which early lesions had intermediate cell numbers and HER2 expression levels. PGR staining intensity was constant, but the number of PGR⁺ cells declined from approximately 40% in normal glands to 0% in primary tumours (Fig. 1d).

The early lesion signature could be activated in 4T1, an MMTV–HER2-negative mouse tumorigenic cell line (Extended Data Fig. 1g) that displays weak endogenous HER2 expression (Extended Data Fig. 1h). Furthermore, HER2-negative mouse tumour cells (67NR and MM3MG cells) did not express the early lesion signature (Extended Data Fig. 1g), but progesterone treatment or transduction of the B isoform of PGR (*Pgr-B*) induced upregulation of HER2 in 4T1 or MM3MG cells, respectively (Extended Data Fig. 1h–i). Collectively, these results suggested that the genetic program of early lesions depends on the combined activation of progesterone and HER2 pathways.

Progesterone induces migration of early lesion cells

Because progesterone mediates branching¹⁴ in mammary gland development, we investigated the role of the progesterone-induced early lesion signature in cancer cell migration. We found that the mRNA levels of the progesterone-induced paracrine signals (PIPS) *Rankl* (also known as *Tnfsf11*) and *Wnt4* were upregulated in early lesion samples (Extended Data Fig. 2a). Treatment of early-lesion-derived cells with PIPS mimicked the effect of progesterone (Extended Data Fig. 2b), suggesting that early lesions exploit the mechanisms of mammary branching for metastasis. Consistent with this, PGR⁺ cells were enriched in distal ducts of normal mammary glands (advancing the branching tree away from the nipple during developmental fat pad invasion) compared to proximal ducts closer to the origin (more differentiated ducts; Extended Data Fig. 2c, d).

Furthermore, progesterone and PIPS induced migration of mammary cells from early-lesion-derived samples (freshly prepared or early-lesion-derived mammospheres) and suppressed migration in cells from primary tumours (Fig. 2a and Extended Data Fig. 2e, f).

cytometric quantification of HER2 and PGR protein expression. Images, representative staining of HER2 (left) and PGR (right). Scale bars, 100 μ m. Mean HER2 staining intensity (red line, left histograms) in arbitrary units (a.u.) and percentage of PGR⁺ cells (right histogram) and box plots. Boxes show lower quartile, median and upper quartile and whiskers indicate minimum and maximum; **** $P \leq 0.0001$ (*t*-test and Stouffer's combined probability test (c) and one-way ANOVA(d)).

PIPS also activate mammary stem cells during mammary gland development¹⁷, prompting us to analyse mammosphere formation¹⁸. Consistent with previous reports on HER2-stimulated stemness^{19,20} BALB–NeuT-derived samples generated significantly higher numbers of spheres than controls (Fig. 2b). Notably, normal mammary cells derived from young (4–9-week-old) mice generated more spheres than cells derived from older mice (Fig. 2b). Furthermore, early-lesion-derived cells generated higher sphere numbers in response to progesterone and PIPS than primary tumour samples ($P \leq 0.01$; Fig. 2c). Neutralizing antibodies against RANKL and a WNT inhibitor abrogated the effect of progesterone on early lesion cell migration and sphere formation (Fig. 2d).

Migrating early lesion cells (that is, those arriving on the other side of the transwell membrane) formed increased sphere numbers in response to progesterone (Fig. 2e and Extended Data Fig. 2g). Oestrogen also induced migration and sphere formation, however, the progesterone inhibitor RU486 inhibited this pro-migratory and pro-sphere-forming effect (Extended Data Fig. 2h), possibly because oestrogen acts via transcriptional induction of *Pgr*²¹. Together, these results suggested that moderate HER2 expression with progesterone or PIPS availability promote sphere-formation and migratory responses in mammary epithelial cells.

HER2 expression levels determine cellular responses

We next performed a series of mechanistic *in vitro* experiments using mouse mammary epithelial MM3MG cells (oestrogen receptor (ER α)-negative but expressing low levels of PGR and HER2). MM3MG cells transduced with *Pgr-B* or *Her2* were subjected to sphere-formation and migration assays (Extended Data Fig. 3). PIPS-responsive cells (in both migration and sphere-formation assays) were HER2^{low}/PGR⁺, whereas PGR⁺ cells themselves were not migrating and HER2^{high} cells were similar to non-migrating primary tumour cells and showed enhanced proliferation.

Cell density regulates HER2 and PGR expression

Notably, individual primary-tumour-derived cells re-expressed *Pgr* mRNA and protein when cultured at low cellular densities (Extended Data Fig. 4a, b). We therefore plated the BALB–NeuT primary-tumour-derived TUBO cell line at different cell densities and found PGR

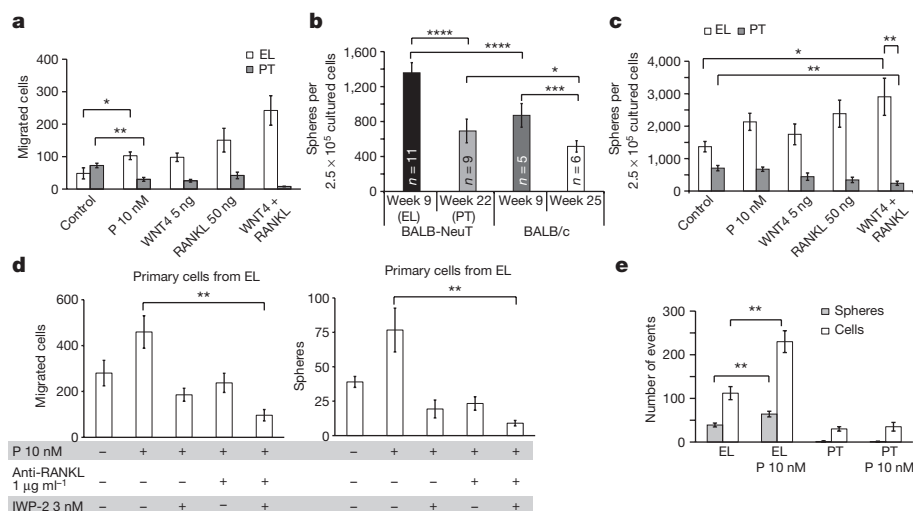


Figure 2 | Progesterone induces migration and sphere formation of early lesion cells. **a**, Early lesion and primary tumour cells respond to progesterone or PIPS (WNT4, RANKL) with increased (early lesions) or decreased (primary tumours) migration. **b**, Mammosphere formation depends on age and HER2 expression. **c**, Cells from early lesions and primary tumours respond to progesterone or PIPS with increased

(early lesions) or decreased (primary tumours) sphere formation. **d**, Depletion of PIPS by IWP-2 (WNT inhibitor) or anti-RANKL (neutralizing antibody) reduces migration (left) and sphere formation (right) of early lesion cells. **e**, Migrating cells activated by PIPS form spheres (see also Extended Data Fig. 2g). Data are mean \pm s.d.; * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$ (Student's *t*-test).

expression in $10 \pm 5\%$ (mean \pm s.d.) of cells grown at low density, but undetectable PGR expression in cells grown at high density (Fig. 3a). Several experiments suggested that there was a soluble factor with PGR-suppressing activity in the vesicular fraction of cell culture supernatants, and that this activity was conserved between mouse

and human (Extended Data Fig. 4b–e). To identify this activity, we analysed the microRNA (miRNA) profiles of supernatants with and without exosomes from TUBO cells and the HER2-overexpressing cell line MM3MG–Her2. miRNA sequencing (Supplementary Data 2) and bioinformatic prediction of *Pgr* regulators (Extended Data Fig. 4f)

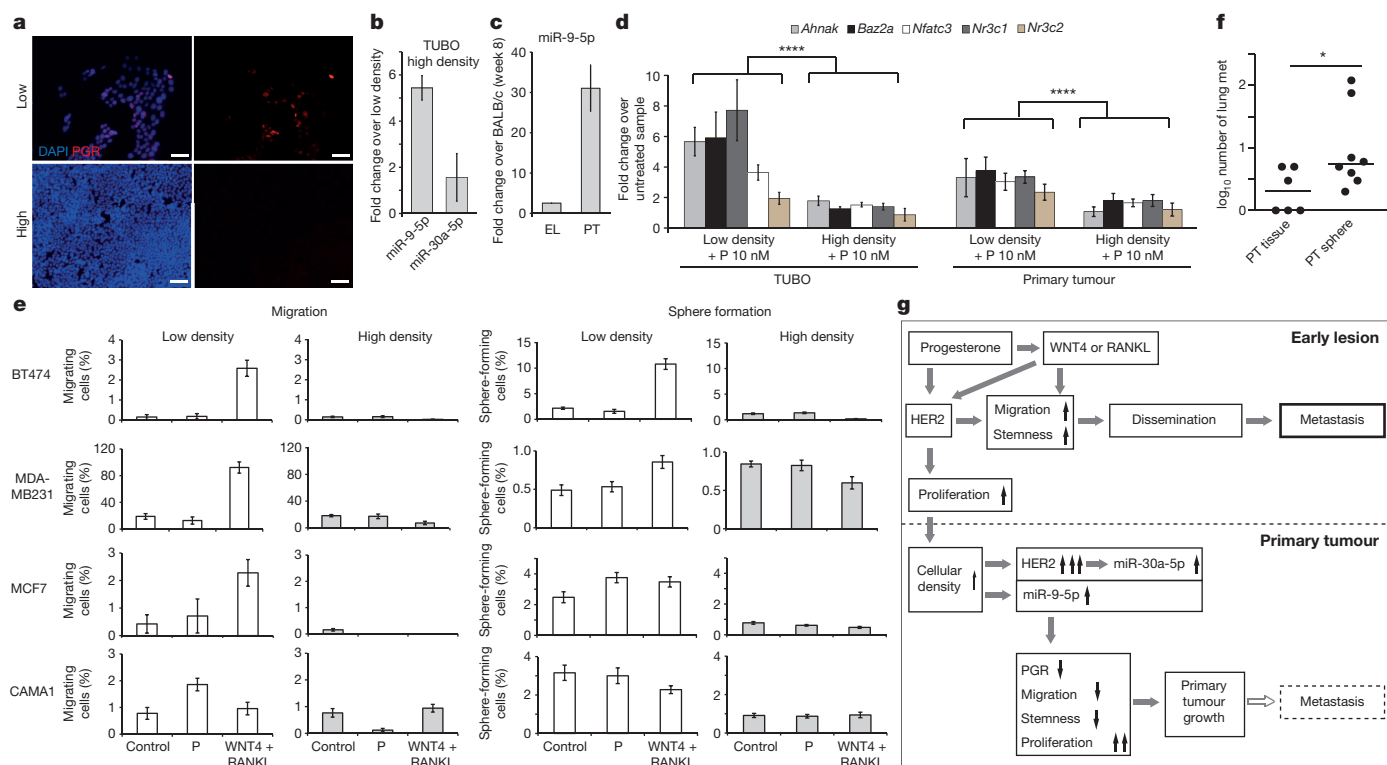


Figure 3 | Cell density regulates PGR expression and early lesion phenotype. **a**, TUBO cells re-express PGR at low cell density. Scale bars, 50 μm . **b**, TUBO cells grown at high density upregulate miR-9-5p. **c**, Expression of miR-9-5p in early lesion and primary tumour samples. **d**, Primary tumour and TUBO cells generate the early lesion signature only when grown at low density. **e**, Migration and sphere-formation of four human cell lines grown at low and high densities and treated with PIPS (WNT4 and RANKL) or progesterone (P) (see also Extended Data Fig. 5).

f, Number of lung macro-metastases (17 weeks after tumour resection) after tumour formation from transplanted tumour pieces (1 mm³; high density) or 50 spheres in 40 μl Matrigel (low density) and primary tumour surgery (shown are median and individual values). **g**, Mechanisms of local tumour and distant metastasis formation as derived from *in vitro* and *in vivo* (see Fig. 4) data. * $P \leq 0.05$; **** $P \leq 0.0001$ (*t*-test and Stouffer's combined probability test (d); mean \pm s.d. (b–e); Mann–Whitney *U*-test (f)).

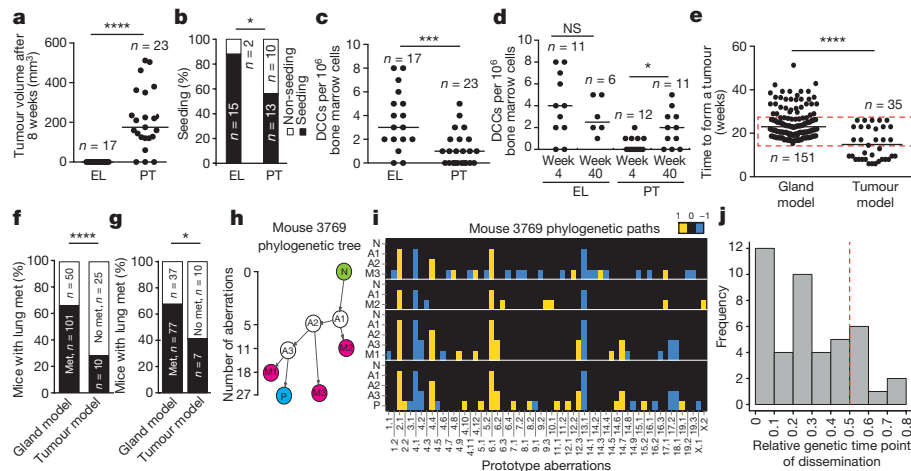


Figure 4 | Progesterone signalling regulates tumour formation and dissemination *in vivo*. **a**, Tumour formation 8 weeks after transplantation of spheres derived from primary tumours or early lesions into mammary fat pads of wild-type BALB/c siblings. **b**, Percentage of mice with DCCs (detected by cytokeratin staining) in bone marrow 8 weeks after transplantation. **c**, Number of DCCs in bone marrow of mice 8 weeks after transplantation. **d**, DCC counts in bone marrow in recipients transplanted at different ages. **e**, Time to tumour formation after transplantation of mammary glands (from 4-week-old mice; gland model) or tumour pieces (from 20–22-week-old mice; tumour model). The red box highlights mice from both models with similar tumour growth kinetics, which are analysed separately in **g** and Extended Data Fig. 6o–q. **f**, Percentage of mice with lung macro-metastasis from gland or primary tumour models. **g**, Macro-metastasis formation in recipient mice with similar tumour

growth kinetics (mice from the red box in **e**; see Extended Data Fig. 6o). **h**, Example of a phylogenetic tree (mouse 3769). A1–A3, inferred common ancestors; M1–M3, metastases 1–3; N, normal cell; P, primary tumour. The ordinate indicates the number of aberrations per profile on a square root scale. **i**, Aberration profiles along tree paths from N via A1–A3 to P or M1–M3 in terms of aberration prototypes (see Extended Data Figs 7–9). **j**, Distribution of relative time points of dissemination on a genetic scale for all 44 primary tumour–metastases pairs. The red line indicates dissemination after which 50% of primary tumour changes were acquired as an arbitrary threshold for early versus late dissemination; see Extended Data Fig. 7d. * $P \leq 0.05$; *** $P \leq 0.001$; **** $P \leq 0.0001$; NS, not significant (Student's *t*-test (**a**, **c**, **d**); Fisher's exact test (**b**); χ^2 test (**f**, **g**); Mann–Whitney *U*-test (**e**)). Lines in **a**, **c**, **d** and **e** denote the median.

identified miR-30a-5p and miR-9-5p as abundantly expressed in TUBO cells and able to downregulate *Pgr* in T47D cells (Extended Data Fig. 4g, h). miR-30a-5p expression was upregulated by *Her2* (Extended Data Fig. 4f, i) and miR-9-5p expression was sensitive to cell density (Fig. 3b, c).

Low cell density induced early-lesion-like features (such as induction of the early lesion signature, and reduced HER2 expression, migration and sphere formation) in TUBO or primary-tumour-derived cells from the BALB-NeuT model (Fig. 3d and Extended Data Fig. 4j–m). Notably, several human breast cancer cell lines shared density regulation of HER2- and PGR-regulating miRNAs (Extended Data Fig. 5a–d). Moreover, we observed density regulation of migration and sphere formation in 10 out of 10 human breast cancer cell lines tested (7 out of 10 by PIPS and 3 out of 10 by progesterone induction; Fig. 3e and Extended Data Fig. 5e).

To assess the contribution of cell density *in vivo*, we compared metastasis formation of transplanted primary tumour pieces (very high cell density) with primary tumourspheres (50 spheres injected in 40 μ l Matrigel, that is, very low density). After primary tumour formation, we performed curative surgery and evaluated metastasis formation. No difference was found in the percentage of mice with metastases (8 out of 19 for tumour pieces compared with 6 out of 16 for spheres), however, animals transplanted with primary tumourspheres had a higher number of metastases ($P \leq 0.05$; Fig. 3f). In summary, we obtained support for a model of metastatic dissemination regulated by cell density, HER2 expression and progesterone signalling (Fig. 3g).

PGR signalling regulates dissemination *in vivo*

To validate these findings, we analysed physiological conditions of reduced (higher age; Extended Data Figs 1f, 2a, 6a, b) and increased (pregnancy) PGR signalling. We transplanted spheres generated from early lesions and primary tumour samples into young (4 weeks of age; PGR-rich) and old (40 weeks of age; PGR-reduced) wild-type recipients. All mice were killed when the first mice developed tumours of 5–10 mm diameter (8 weeks later). Most mice transplanted

with primary tumours had palpable tumours, but none of the mice transplanted with early lesions (Fig. 4a), although these nevertheless harboured microscopic early lesions and/or ductal carcinoma *in situ* (DCIS) (Extended Data Fig. 6c, d). Notably, transplantation of spheres from early lesions increased the number of animals containing DCCs, and resulted in higher numbers of DCCs in bone marrow compared to transplantation of primary tumourspheres (Fig. 4b, c).

There was suppressed dissemination and stimulated tumour formation from primary tumourspheres in the PGR-rich microenvironment of young mammary glands (Fig. 4d and Extended Data Fig. 6e) as expected. Dissemination from early lesions was not reduced in old recipients (Fig. 4d). However, the transplanted early lesions generated a PGR-rich microenvironment in old recipients (Extended Data Fig. 6d), consistent with the observation that cells from early lesions, but not from primary tumours, could generate PGR-positive cells in 3D culture (Extended Data Fig. 6f, g). Co-transplantation of BALB-NeuT primary tumourspheres with MM3MG–Pgr-B into 40-week-old mice resulted in reduction of dissemination and stimulation of tumour formation (Extended Data Fig. 6h, i).

Because progesterone levels are physiologically increased 10- to 20-fold during pregnancy, we mated female transgenic mice at the age when early lesions (week 7) and early tumour formation (week 15) occurred. Mice were killed at term, and those mated at the early lesion age displayed higher numbers of DCCs (Extended Data Fig. 6j); whereas those mated at week 15 formed large tumours within 3 weeks, faster than unmated controls (Extended Data Fig. 6k).

To assess metastasis from early lesions and primary tumour lesions, we transplanted pieces of mammary glands (gland model) of 4–5-week-old transgenic mice or from primary tumours (primary tumour model) into the cleared mammary fat pad of 4-week-old wild-type recipients (Extended Data Fig. 6l, m). Tumour growth to 5–10 mm was fast in animals transplanted with primary tumour pieces (indicating their high viability) and took longer in animals in the gland model (Fig. 4e). After surgical removal of primary tumours (Extended Data Fig. 6l, m), more mice displayed metastasis in the gland model (Fig. 4f; $P < 0.0001$),

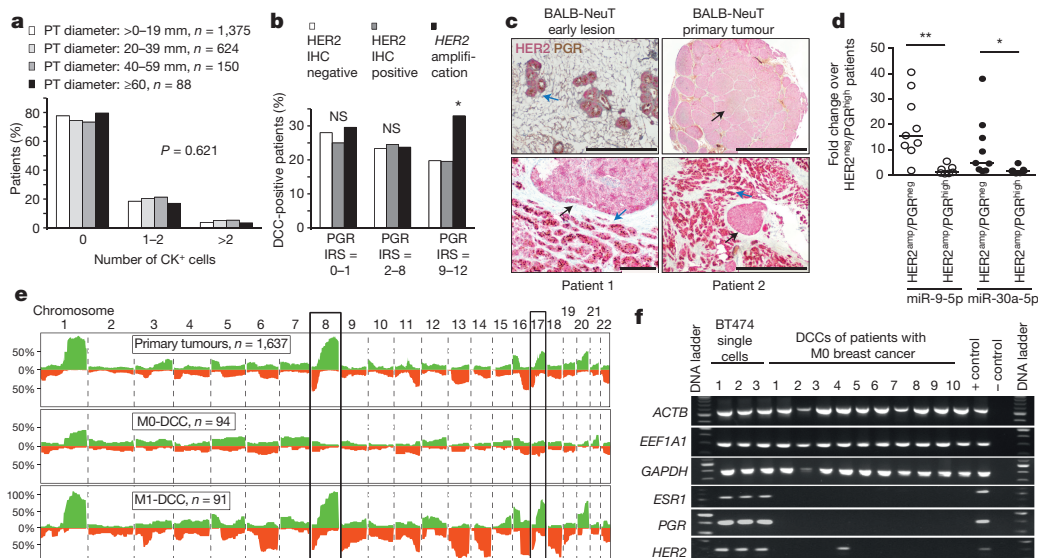


Figure 5 | PGR and HER2 signalling, and dissemination in patients with breast cancer. **a**, An increase in tumour diameter is not accompanied by an increase in DCCs in bone marrow. **b**, PGR and HER2 expression identifies a $HER2^{high}/PGR^{high}$ subgroup of patients with the highest seeding rates. **c**, Comparison of $HER2^{high}/PGR^{high}$ human breast cancers and primary BALB-NeuT mouse tumours for HER2/PGR staining. Top, representative images of an early lesion (left) and primary tumour (right) from the BALB-NeuT model. Bottom, representative images of two patients. High-density regions (strong HER2 and low PGR expression, black arrows) and regions of invasive cells (strong PGR and HER2 expression, blue arrows) are shown. Scale bars, 1 mm (top, bottom right) and 200 μ m (bottom left). **d**, PGR-downregulating miRNAs are repressed

although numbers of metastatic foci were similar in both cases (Extended Data Fig. 6n). Since the slower growth of early lesions until surgery might have extended the available time for metastasis formation, assuming dissemination occurred early, we restricted our analysis to mice with similar kinetics of primary tumour growth (red box in Fig. 4e and Extended Data Fig. 6o). Again, more mice developed lung metastases in the gland model (Fig. 4g), although in the analysis of this subgroup, follow-up time after tumour surgery was significantly longer for mice from the primary tumour model (Extended Data Fig. 6p). The number of metastases was similar (Extended Data Fig. 6q). Together, the *in vivo* results are in line with *in vitro* findings that PIPS induce migration and stemness of early lesion cells and proliferation of advanced primary tumour cells.

We used the gland model to determine which metastases were derived from early rather than late DCCs and performed phylogenetic analyses of 28 primary tumours with 1 or more lung metastases in the same mouse. Phylogenetic trees were generated from copy-number alterations, since BALB-NeuT tumours rarely display point mutations, similar to human breast cancer²². Total numbers of copy-number alterations were indistinguishable between primary tumours and metastases (Extended Data Fig. 7a) and no individual change was significantly associated with primary tumour or metastasis origin (Extended Data Fig. 7b). This suggested that aberrations shared between primary tumours and metastases were acquired earlier, indicating the time of genetic divergence (Fig. 4h, i). In all cases, we observed branching evolution with none or one (single metastasis) or several (multiple metastases; Fig. 4i and Extended Data Figs 7–9) ancestors. To assign derivation of primary tumours and metastases as from early or late DCCs, we assessed the proportion of primary tumour alterations that were already present in the last common ancestor. In linear progression, this value would be 1 (the primary tumour had acquired all mutations before the metastatic precursor left the site). Here, we used the very conservative threshold of ≥ 0.5 for late dissemination. Notably, we found that for 35 out of

in $HER2^{high}/PGR^{high}$ human mammary carcinomas. **e**, Copy-number alterations in human primary breast cancers (from Progenex database) and DCCs isolated from bone marrow of patients with breast cancer with and without metastasis (M0, $n = 94$; M1, $n = 91$). The y axis depicts the percentage of samples with aberrations (green, gain; red, loss) for each chromosomal region. **f**, Oestrogen receptor (*ESR1*) and progesterone receptor (*PGR*) transcript expression in human breast cancer DCCs (10 out of 26 DCCs from 19 M0 patients are shown; see Supplementary Table 8). *ACTB*, *EEF1A1* and *GAPDH* denote controls for sample quality. BT474 single cells are a positive control. * $P \leq 0.05$; ** $P \leq 0.01$; NS, not significant (χ^2 test (a and b); Mann–Whitney *U*-test (d)).

44 individual primary tumour–metastasis pairs (79.5%), lung metastases were derived from eDCCs that disseminated before the primary tumour had acquired 50% of its alterations (Fig. 4j).

HER2 and PGR cooperation in human metastasis

The mechanisms of breast cancer dissemination described above cannot be studied directly in patients, as the event occurs before diagnosis. We therefore checked whether human breast cancers could also seed relatively fewer bone-marrow DCCs (detected by cytokeratin staining²³) with growing primary tumour size, as seen in the BALB-NeuT model (Extended Data Fig. 1a). Indeed, the percentage of DCC-positive patients or DCC numbers in bone marrow of 2,239 patients with breast cancer did not increase with tumour diameter (Fig. 5a). We then investigated whether DCC numbers were associated with HER2 and PGR expression and categorized primary tumours according to their expression levels of PGR and HER2 (Supplementary Table 4). Notably, in breast cancers with a high PGR score, genetic activation of *HER2* increased the dissemination rate ($P \leq 0.05$; Fig. 5b), akin to early lesions in the BALB-NeuT model. This subgroup of patients ($HER2^{amp}/PGR^{high}$) comprised 3.7% of all patients (85 out of 2,239) or 24.6% of patients with *HER2*-amplified tumours (85 out of 345). $HER2^{amp}/PGR^{high}$ (Supplementary Table 5) primary tumours contained *HER2* and *PGR* single- and double-positive and double-negative cells (Extended Data Fig. 10a). Areas of high cell density lacked PGR expression and invasive regions of lower density contained strongly double-positive cells (Fig. 5c), suggesting density-mediated *PGR* regulation within the same samples. We therefore analysed the expression of *PGR*-regulatory miRNAs identified in the BALB-NeuT mouse model (Supplementary Table 6). $HER2^{amp}/PGR^{high}$ tumours displayed lower levels of *PGR*-downregulating miRNAs compared to $HER2^{amp}/PGR^{neg}$ tumours (Fig. 5d), similar to the human $HER2^{high}/PGR^{high}$ cell lines BT474 and T47D (Extended Data Fig. 5a, b). PGR-negative, high-density regions from $HER2^{amp}/PGR^{high}$ samples (Fig. 5c) displayed strong *PGR*-regulating miRNA overexpression (Extended Data Fig. 10b).

DCCs from patients with breast cancer without metastasis (stage M0; $n = 94$ cells; Supplementary Table 7) and with metastasis (stage M1; $n = 91$ cells) were compared to bulk primary breast cancers ($n = 1,637$) from a large database (Progenetix database; <http://www.progenetix.net>). Comparative genome hybridization profiles of primary tumours were very similar to those of single DCCs isolated from M1-stage patients (Fig. 5e). By contrast, single DCCs from M0-stage patients, although these displayed clearly aberrant profiles, lacked the chromosomal gains and losses characteristic of primary tumours, such as 8p loss or 8q gain (Fig. 5e). Thus we can conclude that: (i) DCCs often disseminate before the acquisition of typical breast cancer copy-number alterations; (ii) tumour cells displaying the typical karyotype of established tumours are rarely found at M0-stage of disease in the bone marrow; (iii) DCCs displaying M1-like genomes must replace DCCs with M0-like genomes to generate metastatic disease, even when the primary tumour was surgically removed.

Finally, we tested whether human DCCs ($n = 26$) isolated from the bone marrow of 18 patients with luminal and 1 patient with triple-negative breast cancer (Supplementary Table 8) lacked PGR expression as predicted by the BALB-NeuT model (Extended Data Fig. 3o). We identified and profiled DCCs by combined genome and transcriptome analyses of single cells²⁴. None of the 26 DCCs displaying genomic aberrations expressed *PGR* (Fig. 5f). Notably, the only DCC expressing *HER2* transcripts originated from one of two patients diagnosed with hormone receptor positive DCIS only.

Discussion

This work provides evidence that mouse and human mammary cancer cells migrate and disseminate from morphologically very early lesions. The mechanism, which is shut down as primary tumours grow and overt lesions develop, consists of three major components: cell density, PGR signalling and HER2 signalling. While the specific molecular details are more likely to be tissue dependent than universal, our proposed mechanism may provide a general framework for the understanding of metastasis with cancer cells undergoing a switch from a dissemination to proliferation mode (Fig. 3g).

Our findings challenge the concept that late-disseminating (that is, shortly before surgery), fully mature cancer cells necessarily have a higher ability to form metastases¹³. Indeed, our genetic analyses showed that 80% of metastases in BALB-NeuT mice were derived from eDCCs. The genomic profiles of human DCCs isolated intraoperatively months to years before metastasis represented early cancer cells, not the predominant primary tumour clones, indicating that eDCCs have yet to acquire critical alterations such as gains on chromosome 8q to form metastases. The time to acquire such changes may largely account for the long latency periods and late relapses, which are becoming clinically more and more relevant²⁵. Therefore, parallel progression⁷ seems to be typical rather than exceptional. This is supported by a recent sequencing study, in which not a single case of linear progression from primary tumour to metastasis was found²⁶.

Our data indicate that breast cancer hijacks a developmental program, in which progesterone and its paracrine signals regulate mammary epithelial branching, fat pad invasion^{14,27,28} and mammary stem-cell expansion^{14,17,18,29–31} during development and pregnancy, for dissemination, independently of breast cancer subtype. Relevance to human disease is highlighted by a careful analysis of mortality from DCIS, hitherto defined as pre-invasive lesion³². Of the 3% of DCIS fatalities, more than 50% die of metastasis without local relapse, indicating lethal dissemination before surgery of the pre-invasive lesion³². Evidently, dissemination will also occur early if the tumour is not diagnosed as DCIS but later as invasive cancer. Moreover, death from DCIS increased significantly to 8% for young women³², possibly because in young women the epithelial compartment and the micro-environment have high PGR expression.

The gradual generation of a HER2^{high}/PGR[−] phenotype may explain why early lesions of the BALB-NeuT model, but not advanced tumours,

were found to represent human luminal tumours³³. The BALB-NeuT model apparently represents a mixed luminal and HER2 phenotype and models several breast cancer subtypes. Cell density, progesterone, PIPS and HER2 signalling regulated dissemination and sphere formation in breast cancer cell lines of all subtypes similar to the BALB-NeuT model. Notably, HER2-positive circulating tumour cells were detected in DCIS and lobular carcinoma *in situ* or patients with M0-stage breast cancer irrespective of the HER2 status of the primary tumour³⁴.

High cell density and HER2 expression were responsible for the proliferative switch of mammary epithelial cells. Oncogenic mutations characterize benign tumours that do not metastasize^{35–37}; indeed, strong activation of oncogenic pathways represses metastasis while increasing proliferation^{38,39}. However, our experiments do not exclude the possibility that metastases form from advanced tumours because the early lesion dissemination program may become re-activated in areas of low cell density.

Our findings have implications for the understanding of metastasis and development of therapies: systemically spread cancer cells probably comprise cells derived from different stages of primary tumour evolution, including the earliest. Since DCCs from early and later stages have metastatic potential, therapies targeting the seed of metastasis need to address this heterogeneity.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 October 2015; accepted 11 November 2016.

Published online 14 December 2016.

1. Aguirre-Ghiso, J. A., Bragado, P. & Sosa, M. S. Metastasis awakening: targeting dormant cancer. *Nat. Med.* **19**, 276–277 (2013).
2. Polzer, B. & Klein, C. A. Metastasis awakening: the challenges of targeting minimal residual cancer. *Nat. Med.* **19**, 274–275 (2013).
3. Cole, B. F., Gelber, R. D., Gelber, S., Coates, A. S. & Goldhirsch, A. Polychemotherapy for early breast cancer: an overview of the randomised clinical trials with quality-adjusted survival analysis. *Lancet* **358**, 277–286 (2001).
4. Gianni, L. *et al.* Treatment with trastuzumab for 1 year after adjuvant chemotherapy in patients with HER2-positive early breast cancer: a 4-year follow-up of a randomised controlled trial. *Lancet Oncol.* **12**, 236–244 (2011).
5. Klein, C. A. *et al.* Genetic heterogeneity of single disseminated tumour cells in minimal residual cancer. *Lancet* **360**, 683–689 (2002).
6. Schmidt-Kittler, O. *et al.* From latent disseminated cells to overt metastasis: genetic analysis of systemic breast cancer progression. *Proc. Natl Acad. Sci. USA* **100**, 7737–7742 (2003).
7. Klein, C. A. Parallel progression of primary tumours and metastases. *Nat. Rev. Cancer* **9**, 302–312 (2009).
8. Hüsemann, Y. *et al.* Systemic spread is an early step in breast cancer. *Cancer Cell* **13**, 58–68 (2008).
9. Rhim, A. D. *et al.* EMT and dissemination precede pancreatic tumor formation. *Cell* **148**, 349–361 (2012).
10. Eyles, J. *et al.* Tumor cells disseminate early, but immunosurveillance limits metastatic outgrowth, in a mouse model of melanoma. *J. Clin. Invest.* **120**, 2030–2039 (2010).
11. Banyas, M. *et al.* Hematogenous and lymphatic tumor cell dissemination may be detected in patients diagnosed with ductal carcinoma *in situ* of the breast. *Breast Cancer Res. Treat.* **131**, 801–808 (2012).
12. Sängster, N. *et al.* Disseminated tumor cells in the bone marrow of patients with ductal carcinoma *in situ*. *Int. J. Cancer* **129**, 2522–2526 (2011).
13. Valastyan, S. & Weinberg, R. A. Tumor metastasis: molecular insights and evolving paradigms. *Cell* **147**, 275–292 (2011).
14. Fernandez-Valdivia, R. *et al.* Revealing progesterone's role in uterine and mammary gland biology: insights from the mouse. *Semin. Reprod. Med.* **23**, 22–37 (2005).
15. Mulac-Jericevic, B., Lydon, J. P., DeMayo, F. J. & Conneely, O. M. Defective mammary gland morphogenesis in mice lacking the progesterone receptor B isoform. *Proc. Natl Acad. Sci. USA* **100**, 9744–9749 (2003).
16. Aupperlee, M. D., Smith, K. T., Kariagina, A. & Haslam, S. Z. Progesterone receptor isoforms A and B: temporal and spatial differences in expression during murine mammary gland development. *Endocrinology* **146**, 3577–3588 (2005).
17. Briskin, C. Progesterone signalling in breast cancer: a neglected hormone coming into the limelight. *Nat. Rev. Cancer* **13**, 385–396 (2013).
18. Liao, M. J. *et al.* Enrichment of a population of mammary gland cells that form mammospheres and have *in vivo* repopulating activity. *Cancer Res.* **67**, 8131–8138 (2007).

19. Korkaya, H., Paulson, A., Iovino, F. & Wicha, M. S. HER2 regulates the mammary stem/progenitor cell population driving tumorigenesis and invasion. *Oncogene* **27**, 6120–6130 (2008).
20. Ginestier, C. *et al.* ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* **1**, 555–567 (2007).
21. Lee, Y. J. & Gorski, J. Estrogen-induced transcription of the progesterone receptor gene does not parallel estrogen receptor occupancy. *Proc. Natl Acad. Sci. USA* **93**, 15180–15184 (1996).
22. Stephens, P. J. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400–404 (2012).
23. Fehm, T. *et al.* A concept for the standardized detection of disseminated tumor cells in bone marrow from patients with primary breast cancer and its clinical implementation. *Cancer* **107**, 885–892 (2006).
24. Gužvić, M. *et al.* Combined genome and transcriptome analysis of single disseminated cancer cells from bone marrow of prostate cancer patients reveals unexpected transcriptomes. *Cancer Res.* **74**, 7383–7394 (2014).
25. Klein, C. A. Framework models of tumor dormancy from patient-derived observations. *Curr. Opin. Genet. Dev.* **21**, 42–49 (2011).
26. Brastianos, P. K. *et al.* Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov.* **5**, 1164–1177 (2015).
27. Briskin, C. *et al.* Essential function of Wnt-4 in mammary gland development downstream of progesterone signaling. *Genes Dev.* **14**, 650–654 (2000).
28. Fernandez-Valdivia, R. *et al.* The RANKL signaling axis is sufficient to elicit ductal side-branching and alveologenesis in the mammary gland of the virgin mouse. *Dev. Biol.* **328**, 127–139 (2009).
29. Asselin-Labat, M. L. *et al.* Steroid hormone receptor status of mouse mammary stem cells. *J. Natl. Cancer Inst.* **98**, 1011–1014 (2006).
30. Joshi, P. A. *et al.* Progesterone induces adult mammary stem cell expansion. *Nature* **465**, 803–807 (2010).
31. Mukherjee, A. *et al.* Targeting RANKL to a specific subset of murine mammary epithelial cells induces ordered branching morphogenesis and alveologenesis in the absence of progesterone receptor expression. *FASEB J.* **24**, 4408–4419 (2010).
32. Narod, S. A., Iqbal, J., Giannakeas, V., Sopik, V. & Sun, P. Breast cancer mortality after a diagnosis of ductal carcinoma *in situ*. *JAMA Oncol.* **1**, 888–896 (2015).
33. Astolfi, A. *et al.* Gene expression analysis of immune-mediated arrest of tumorigenesis in a transgenic mouse model of HER-2/neu-positive basal-like mammary carcinoma. *Am. J. Pathol.* **166**, 1205–1216 (2005).
34. Ignatiadis, M. *et al.* HER2-positive circulating tumor cells in breast cancer. *PLoS One* **6**, e15624 (2011).
35. Hafner, C. *et al.* Oncogenic PIK3CA mutations occur in epidermal nevi and seborrheic keratoses with a characteristic mutation pattern. *Proc. Natl Acad. Sci. USA* **104**, 13450–13454 (2007).
36. Hafner, C. *et al.* Multiple oncogenic mutations and clonal relationship in spatially distinct benign human epidermal tumors. *Proc. Natl Acad. Sci. USA* **107**, 20780–20785 (2010).
37. Pollock, P. M. *et al.* High frequency of BRAF mutations in nevi. *Nat. Genet.* **33**, 19–20 (2003).
38. Hutchinson, J. N., Jin, J., Cardiff, R. D., Woodgett, J. R. & Muller, W. J. Activation of Akt-1 (PKB- α) can accelerate ErbB-2-mediated mammary tumorigenesis but suppresses tumor invasion. *Cancer Res.* **64**, 3171–3178 (2004).
39. Liu, H. *et al.* MYC suppresses cancer metastasis by direct transcriptional silencing of α_V and β_3 integrin subunits. *Nat. Cell Biol.* **14**, 567–574 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. Perry for his critical reading of the manuscript. This work was supported by grants from the DFG (KI 1233/2-1, KL 1233/3-1, KL 1233/10-1 (C.A.K.); Me2064/4-1, (G.M.); SP 938/2-1 (R.S.); INST 89/341-1 FUGG (TissueFAX)); the Dr Josef Steiner Foundation and ERC (322602) to C.A.K.; the SWCRF, CA109182, CA196521, CA163131, BC132674 (J.A.A.-G.) F31 CA183185 (K.L.H.) BC112380 (M.S.S.), NIH 1S10RR024745 Microscopy CoRE at ISMMS and by a donation from A. Jungmayer.

Author Contributions C.A.K. and H.H. designed and evaluated core experiments. H.H. performed core experiments. M.M.S.O., C.W., L.K.N., C.E., C.R. and M.Gu. helped with *in vivo*, *in vitro* and primary culture experiments. M.H., M.M., M.Gr., R.S. and H.H. performed bioinformatic and statistical analyses. M.W.-K., K.L.H., M.S.S. and F.W. performed staining and pathological analysis. N.E. and G.M. performed miRNA sequencing and analysis. G.H., N.P., A.K.H., F.-A.T., S.Y.B., B.R., S.B. and T.F. performed DCC analysis or collected patient data. C.A.K. and H.H. wrote the manuscript with input from J.A.A.-G.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.A.K. (christoph.klein@ukr.de).

METHODS

Mice. BALB-NeuT transgenic mice were obtained through collaboration with G. Forni and maintained in our facilities according to the European Union guidelines. All animal experiments were performed according to the EU and national institutional regulations. Mice were screened at 3–4 weeks of age for hemizygosity (neuT⁺/neuT⁻), and negative littermates served as wild-type BALB/c mice controls. Mammary glands of BALB-NeuT female mice were inspected twice a week and arising tumours were measured in two perpendicular diameters. Data acquisition for bone-marrow DCCs was performed in a blinded manner, whereas enumeration of lung metastasis was performed unblinded by two observers. All experimental animal procedures were approved and conducted according to German federal and state regulations (Government of Upper Palatinate, 55.2-2532.1-27/14).

Surgery and transplantation experiments. Mice were anaesthetized with midazolam 5 mg kg⁻¹, fentanyl 0.05 mg kg⁻¹, medetomidin 0.5 mg kg⁻¹ by intraperitoneal injection. The thorax and abdomen were shaved; the skin was incised caudal to cranial along the midline. Fifty spheres were mixed with Matrigel (BD Biosciences: 356231, final concentration, 40%) and injected into the fourth right mammary gland of BALB/c mice (4 and 40-week-old mice). For tissue transplantation, a piece (approximately 1 mm³) of donor mammary tissue from 4-week-old BALB-NeuT mice (gland model) or primary tumours (primary tumour model) was implanted in the cleared mammary fat pad of recipient mice (4-week-old BALB/c mice). The skin was closed by a suture using polygelatin string (Ethicon) and anaesthesia was antagonized with flumazenil 0.5 mg kg⁻¹, atipamezol 2.5 mg kg⁻¹, naloxon 1.2 mg kg⁻¹ by subcutaneous injection. Postoperative analgesia was achieved by buprenorphin (0.1 mg kg⁻¹) by subcutaneous injection. Curative surgery or dissection was done when the diameter of tumours was between 5–10 mm. After surgery, mice were kept until we observed the first general signs of reduced health. After dissection, lungs were macroscopically inspected and individual metastases counted.

Extent of dissemination relative to tumour area. Data were taken from ref. 8. Briefly, gland or tumour areas were calculated from 270 mammary glands or tumours of 27 mice assuming the shape of an ellipse or circle for each tumour. The tumour area of mammary glands without palpable (that is, not measurable by a caliper) tumours was set to 0.1 mm² (that is, assuming a diameter of 350 µm of a total, circular hyperplastic lesion within a mammary gland) for lesions from 4–9-week-old mice and 0.4 mm² for 11-week-old mice. The adjustment for 11-week-old mice was based on a microscopic evaluation showing an about 4-fold increase in hyperplastic lesions. Dissemination to the bone marrow was determined by the number of cytokeratin-positive cells per 10⁶ bone-marrow cells (Extended Data Fig. 1a).

Mouse bone marrow preparation and staining for DCCs. Bone marrow was collected from femurs and tibiae. The bone marrow was rinsed with a 26-G needle with 1 ml of PBS. After density gradient centrifugation, 5 × 10⁵ interphase cells were put on adhesion slides (Menzel). At least 10⁶ cells per mouse were stained to detect positive cells. Blocking solution (5% rabbit serum in 1 × TBS (50 mM Tris-base, 150 mM NaCl, pH 7.4)) was added to the slides to rehydrate the cells and to block unspecific binding of antibodies to the cells. After 20 min the blocking solution was discarded and primary antibody against cytokeratins 8 and 18 (CK8/18; all antibodies and working concentrations are in Supplementary Table 9) or guinea pig serum (the CK antibody originated from guinea pig) as control, was added and slides were incubated for 60 min. The primary antibody was discarded and the slides were washed 3 × for 3 min in 1 × TBS. The slides were incubated with the secondary antibody for 25 min, and then washed 3 × for 3 min in 1 × TBS followed by incubation with the ABC complex (Vector Laboratory) for 25 min. Finally, the development system of the BCIP/NBT (AP Conjugate Substrate Kit, Bio-Rad Laboratories GmbH, 1706432; Levamisol hydrochloride, Sigma-Aldrich GmbH, L-9756) for alkaline phosphatase enzymatic substrate was added for 10 min. The slides were washed 3 × for 3 min and screened for CK8/18-positive cells. The positive cells were typically violet-to-black in colour. TUBO, a tumour cell line derived from a mouse primary mammary tumour of BALB-NeuT and known to express CK8/18, was used as a positive control.

Laser microdissection and microarray analysis. Laser microdissection (PALM MicroBeam from Carl Zeiss MicroImaging GmbH) was performed to dissect metastatic lesions from lungs, primary tumours and epithelial layers of mammary glands of BALB-NeuT mice at the time point of early lesions (7–9-week-old mice; examples are shown in Extended Data Fig. 1b), and BALB/c mice at different ages (description of samples is given in Supplementary Table 1). Small pieces adding up to 100,000 µm² for each sample were catapulted into a cap with 10 µl paramagnetic, biotinylated, oligo-dT-peptide, nucleic-acid, bead suspension and lysis buffer (Active Motif, 29011). Extraction of mRNA and microarray experiments were performed as described previously⁴⁰. Heatmaps in Fig. 1a were generated using Euclidean distance and complete linkage agglomerative clustering on row (gene)-wise standardized expression data (zero mean, unit standard deviation).

Cell lines, cell culture and cell stimulation. Breast cancer cell lines (4T1, 66cl4, and 67NR) were provided by F. Miller. These cell lines were derived from a single mammary tumour that arose spontaneously in a wild-type BALB/cF3H mouse. The MM3MG mouse mammary epithelial cell line derived from a BALB/c background was purchased from ATCC (ATCC CRL6376). The TUBO cell line is a cloned cell line established *in vitro* from a lobular carcinoma that arose spontaneously in a BALB-NeuT mouse (gift from G. Forni). All mouse and stably transduced cell lines were grown in DMEM medium (Pan-Biotech, P04-03500) supplemented with 10% (20% for TUBO cell line) FCS (Pan-Biotech: P30-3702), 2 mM L-glutamine (Pan-Biotech, P04-80100), 10 U ml⁻¹ penicillin/streptomycin (Pan-Biotech, P06-07050). All human cell lines were purchased from ATCC and each cell line was maintained in medium recommended by ATCC. The origin of the cell lines was confirmed by short tandem repeat (STR) analysis (Cell-ID, Promega). All cells were incubated at 37 °C with 5% CO₂. Steroid hormones (progesterone, aldosterone, β-oestradiol, testosterone and hydrocortisone; all from Sigma-Aldrich) and RU486 (Sigma-Aldrich) were dissolved in ethanol. RANKL (mouse Rankl, Abcam, ab151200; human RANKL, Abcam, ab9958); WNT4 (mouse WNT4, R&D systems, 475-WN; human WNT4, Abnova, H00054361-P01); lapatinib (Santa Cruz Biotechnology, SC202205); IWP-2 (Sigma-Aldrich, I0536); RANKL-neutralizing antibody (Lifespan Biotech, LS-C150261) were dissolved according to the manufacturer's instructions. All cell lines were routinely tested for mycoplasma and were found to be negative.

Primary cultures and sphere formation assay. Fresh mammary glands or primary tumours were digested with 200 units per ml collagenase I (Worthington Biotech, L5004196) and 1 µg ml⁻¹ hyaluronidase (Sigma-Aldrich, 4272) in basal medium for 2 h at 37 °C. The basal medium consisted of DMEM/F12 (PAN biotech, P04-41450) supplemented with 10 mM HEPES buffer (Sigma-Aldrich, H0887), penicillin/streptomycin (Pan Biotech, P1-010) and 10 µg ml⁻¹ insulin (Sigma-Aldrich, I9278). Digested tissue cells were centrifuged and re-suspended in basal medium. The cells were subsequently cultured at a density of 5 × 10⁴ cells per ml in ultra-low adherent plates coated with 1.2% poly-HEMA (Sigma-Aldrich, P3932) or at a density of 2.6 × 10⁴ cells per cm² for adherent culture in DMEM medium (Pan-Biotech, P04-03500) supplemented with 10% FCS (Pan-Biotech, P30-3702), 2 mM L-glutamine (Pan-Biotech, P04-80100), 10 U ml⁻¹ penicillin/streptomycin (Pan-Biotech, P06-07050). Sphere culture medium was basal medium supplemented with 2% B27 (Gibco, 17504044), 10 µg ml⁻¹ EGF (Sigma-Aldrich, E9644), 10 ng ml⁻¹ bFGF (Sigma-Aldrich, F0291), 20 ng ml⁻¹ hIL6 (gift from S. Rose-John), 4 ng ml⁻¹ heparin (Sigma-Aldrich, H3149), 5 ng ml⁻¹ GRO-α (R&D systems, 275-GR). Concentrations of activators, inhibitors and other molecules are given in the main text, figures or legends. Sphere cultures were incubated at 37 °C with 5% CO₂ and 7% O₂ and cultures were screened for spheres after 10 days. Only spheres with a diameter over 50 µm were counted. The size of mammospheres was inspected under a light microscope and measured using Zeiss Axiovision software (Carl Zeiss) after 10 days.

Cell density experiments. TUBO cells were cultured at 3 × 10⁴ cells per cm² for high density and 5.2 × 10³ cells per cm² for low-density experiments. Primary cells derived from primary tumours were cultured in 10.6 × 10⁴ cells per cm² for high-density and 2.2 × 10³ cells per cm² for low-density experiments. Density criteria for human cell lines were 100% confluency for high density and 20–30% confluency for low-density. For hormone treatment and comparisons between low and high-density experiments, cells were incubated for 76 h with fresh hormone treatment and washes (2 × with PBS) at 24-h intervals. We avoided changing medium and washing during incubation of cells for miRNA analyses. In migration experiments we seeded 10⁴ cells per well (24-well migration chambers) for low-density experiments for all cell lines and 5 × 10⁴ cells per well for high-density experiments with TUBO cells and 4 × 10⁴ for the other cell lines.

Transwell assay. Transwell inserts (Corning, 3419) with a microporous membrane of 0.4 µm were used to separate the upper and lower compartments. The microporous membrane allows only soluble factors to pass between the compartments. Early lesion cells were cultured in the lower chamber and primary tumour cells were cultured in the upper chamber. Both were cultured at a density of 10⁶ cells per well of 6-well plates (DMEM with 10% FCS).

Migration/sphere-formation assays. Transwell inserts (Corning: 3422) with 8-µm pores were coated with 30% matrigel. 4 × 10⁴ from cell lines and 10⁵ cells isolated from tissue or dissociated spheres were resuspended in FCS-free medium (DMEM) before seeding. Cells were then seeded in 200 µl of FCS-free medium on top of the Matrigel layer and FCS medium (DMEM containing FCS) was added to the lower chamber. For additional treatments medium in both upper and lower compartments was supplemented with the reagents at concentrations specified in the text and figures. After incubating cells isolated from mammospheres or from freshly digested tissues for 72 h, inserts were removed and cells were fixed with methanol (−20 °C for 10 min) and stained with trypan blue. Cells were counted from 3 fields (4 × magnification) when visualized under the microscope.

For the combined migration/sphere formation assay, cells were placed on a layer of 30% Matrigel in the upper chamber and the lower chamber was coated with poly-HEMA. The mammosphere medium used is described above. After 72 h, inserts were removed, fixed and stained with trypan blue for single cell migration analysis. 600 µl fresh sphere medium was added to the lower chamber and cells were incubated for 11 days when spheres were counted (Fig. 2e and Extended Data Fig. 2g).

Proliferation assay. Single-cell suspensions were cultured in 96-well plates (Corning Inc) and proliferation was evaluated by a XTT-colorimetric assay kit (Roche, 11465015001) based on the manufacturer's instructions. Seeding concentration of cells was 3,000 cells per well. The experiment was performed with 6 technical replicates. The medium was supplemented with the tested factors or hormones and vehicle (see corresponding experiments) and was changed every second day.

Immunohistochemistry. For PGR and HER2 immunohistochemistry of tissue sections, we used 5-µm sections of paraffin blocks placed onto poly-L-lysine-coated slides. Samples were dewaxed by two 5-min washes in xylene and rehydrated with graded alcohol by 5-min washes and a final wash in water. A standard Tris-EDTA buffer and pressure-cooking was used for antigen retrieval and then sections were blocked in 0.3% H₂O₂ in TBS and 10% normal goat serum. Sections were incubated for 1 h with primary antibodies and, after washing, secondary antibodies (Vector laboratory, PK4001 or PK5000) were added based on the manufacturer's recommended dilution (see Supplementary Table 9). After washing with PBS, sections were stained using the ABC detection system (Vector Laboratory) according to the manufacturer's instruction. Visualization was performed with chromogen reagent (Dako, 10046560) according to manufacturer's instructions.

Immunofluorescence staining. For staining of cells from monolayer cell cultures, cells were seeded onto 24-well culture plates at an appropriate density. After 72 h of incubation, cells were washed with PBS and fixed with 4% PFA for 10 min. Then, cells were permeabilized with 0.2% Triton X-100 followed by washing steps and blocking with 1% BSA in PBS at 37 °C and incubated with primary antibody (see Supplementary Table 9) for 1 h at room temperature. Cells were then washed 3 × with PBS and incubated with labelled secondary antibody (Jackson ImmunoResearch Laboratory Inc) for 1 h at room temperature. For nuclear counterstaining, cells were incubated for 10 min with 0.5 µg ml⁻¹ DAPI (Sigma-Aldrich). For the staining of spheres in differentiation experiments, mammospheres were picked and transferred to a 24-well cell-culture plate and incubated for 8 h in sphere medium in order to fix them to the surface. The subsequent staining protocol was as for monolayer cell culture staining. For staining of cells attached to the inserts from migration experiments, inserts were used directly after migration (see migration assay), for the blocking step and immunofluorescence staining, the monolayer cell culture staining procedure was applied. Images were captured on an AxioVert 200M microscope (Carl Zeiss Microscopy).

Quantification of HER2 and PGR staining in tissue sections by TissueFAX cytometry. Tissue sections were stained with an automated staining machine (Ventana, BenchMark ULTRA). Tissue sections used for analysis were stained within the same run. Images of stained tissue sections were scanned with the TissueFAXSi-plus imaging system (TissueGnostics, Vienna, Austria; acquisition software: TissueFAXS v3.5.129) equipped with a digital Pixelink colour camera (PCO AG). Images for the analysis of HER2 and PGR staining were analysed with HistoQuest software v3.5.3.0185 (TissueGnostics). Using the HistoQuest software, two markers were created: haematoxylin as master marker (nucleus) and HER2 or PGR as non-master marker. To achieve optimal cell detection, the following parameters were adjusted: (i) nuclei size; (ii) discrimination by area; (iii) discrimination by grey and (iv) background threshold. For the evaluation of the HER2 staining intensity of cells or the percentage of PGR-expressing cells, histograms were created, allowing the visualization of corresponding cells in the source region of interest using the real-time back-gating feature. The cut-off discriminated between false events and specific signals according to cell size and intensity of staining. For HER2 staining, 38,675 primary tumour cells (6 regions, 1.99 mm²), 28,850 cells from hyperplastic regions (25 regions, 1.55 mm²) and 14,938 cells from non-transformed ducts (30 regions, 0.93 mm²) were analysed. For PGR, 12,269 cells of early lesions (hyperplasia, 7 weeks, 11 regions, 0.5 mm²), 12,702 cells of non-transformed (normal duct, 7 weeks, 56 regions, 0.7 mm²) and 25,357 primary tumour cells (9 regions, 1.3 mm²) were analysed (Fig. 1d).

Quantitative PCR. All mRNA extractions were performed using the RNeasy kit (Qiagen, 74104) according to the manufacturer's instructions. For miRNA extraction, the miScript II RT Kit (Qiagen, 217004) was used. cDNA was generated using a reverse transcriptase kit (Qiagen, 205311 for total RNA and 218161 for miRNA). Finally, 25 ng of cDNA was used for qPCR. qPCR was performed using a LightCycler instrument (Roche) and Fast Start Master SYBR Green Kits (Roche). Data analysis was done using the RelQuant software (Roche) with a reference gene and a calibrator (reference) sample in every run. Mouse reference cDNA

served as a positive control. Samples with unspecific products in the melting curve analysis were discarded from further analysis. Expression levels are given relative to *Actb* (β-actin) for gene expression analyses and *Rnu6* for miRNA analyses (primer sequences are provided in Supplementary Table 10). All primers for mRNA analyses were synthesized by Eurofins MWG Operon, and by Qiagen for miRNA analyses.

For comparison of miRNA levels in high-density and low-density regions (Extended Data Fig. 10b) of formalin-fixed paraffin-embedded samples, regions were punched out using a 1.5-mm puncher (PFM medical; 48115). Samples were incubated for 10 min at 70 °C followed by xylene-ethanol de-paraffinization and overnight proteinase K (0.5 µg µl⁻¹, Roche 03115828001) digestion. Then miRNA extraction was performed using the miRNeasy kit. For comparison of miR-30a-5p and miR-9-5p between HER2^{high}/PGR^{high} human mammary carcinomas and HER2^{high}/PGR⁻ carcinomas, miRNAs were extracted from freshly frozen samples using the miRNeasy kit. Expression of miR-9-5p and miR-30a-5p was normalized to HER2⁻/PGR^{high} breast cancers (see Supplementary Table 6 for details on patients).

Lenti- and gamma retroviral transduction. PGR expression was carried out with a lentiviral construct encoding human *PGR-B* (GeneCopeia, Z5911). Lentiviral packaging was conducted as previously described⁴¹. Helper vectors were pSPAX2 and pMD2.G (Addgene). Selection was performed using 10 µg ml⁻¹ of puromycin (Sigma-Aldrich, P8833). For *Her2* expression pLXSN-NNeu (rat wild-type *Neu/Her2*) was used (obtained from L. Pettit)⁴². Retroviral delivery of transgenes was performed as described previously⁴³. Helper vectors were pCMV-VSV-G and pUMVC3 obtained from Addgene. Selection was performed using 1,000 µg ml⁻¹ of G418 (Sigma-Aldrich, G9516). MM3MG cells were transduced with lentiviruses and/or retroviral vectors and cell colonies were selected using antibiotics. Positively transduced clones were expanded and screened for PGR and/or HER2 levels by western blot analysis and qPCR.

Western blot analysis. Cell lysates were prepared using RIPA buffer (Sigma-Aldrich, R0278) and were analysed with the BCA protein assay kit (Thermo Scientific, 23227) to measure and their protein concentration was equalized. Quantified protein lysates were resolved on 6.5% SDS-PAGE gels, and transferred onto a polyvinylidene difluoride membrane (Millipore), and immunoblotted with the primary antibodies overnight followed by incubation with the horseradish peroxidase-conjugated secondary antibodies. The blots were visualized using a substrate kit (GE Healthcare, RPN2109) and bands were visualized by Imagequant LAS 4000 (GE Healthcare). The full blot images are shown in Supplementary Fig. 1.

Exosome isolation, miRNA analyses and sequencing. To prepare conditioned medium, TUBO cells were seeded at a density of 3 × 10⁴ cells per cm². After 4 days, medium was collected, centrifuged and filtered and used as conditioned medium. For exosome isolation we used an ultracentrifugation method as previously described⁴⁴. Exosome pellets were resuspended in fresh medium and used for T47D cell line treatment. PGR expression was checked at different time points (4, 8, 24, and 48 h). For miRNA sequencing we used 4 × 10⁶ cells and exosomes were isolated from confluent medium from TUBO cells. The miRNA cloning and sequencing was done as described previously⁴⁵. All pooled samples were sequenced on a MiSeq system (Illumina) in a single-end run with 80 cycles using the MiSeq reagent kit v3. Data analysis was performed using in-house written scripts. Sequences were mapped—without any mismatches allowed—against mouse miRNAs listed in the miRBase v20 (June 2013; <http://www.mirbase.org>). The minimum length of reads was set to 18 nucleotides. Annotated miRNA-reads were normalized as RPM values according to the total number of mapped reads in the respective library. Mimic miRNAs were ordered from Eurofins MWG Operon Company and all sequences are listed in Supplementary Table 11. For miRNA transfection we used reverse transfection protocols according to instructions for RNAiMAX (Life Technology, 13778030) and with a 50 nM concentration of miRNA.

Mouse genome comparative genome hybridization microarrays. DNA samples were extracted from freshly frozen samples using the DNeasy Blood & Tissue Kit (Qiagen, 69504). Genomic DNA labelling was done using the Agilent SureTag DNA Labelling Kit (Agilent, 5190-3400). Array CGH was performed on oligonucleotide-based SurePrint G3 Mouse CGH Microarray Kit, 4x180K (design code: 027411) according to the protocol provided by the manufacturer (Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis, v.7.2, July 2012).

Phylogeny of primary tumours and metastases. Ancestral relations among matched samples of primary tumour and metastases were inferred using array CGH profiles. The array comparative genomic hybridization dataset consisted of 28 primary tumour samples with 1–3 corresponding metastasis samples (18 primary tumours with 1 metastasis; 4 with 2 metastases and 6 with 3 metastases).

Positions of the probes on the array were mapped to the current mouse reference genome (mm10) using the liftOver tool⁴⁶. No background correction was applied to the data⁴⁷. The data were first normalized within arrays using Loess⁴⁸. Then, log ratios were corrected for spatial artefacts using a median filter with an

11 × 11 block of probes⁴⁹ and a between array scale normalization was applied⁵⁰. Duplicate probes (having the same genomic position) were summarized by their median log ratio. The R package limma⁵¹ (v.3.28.1) was used for normalization within arrays and between arrays (with default parameters). In a final step, wavy patterns were removed from the data using an approach similar to ref. 52, but with modifications to account for broad copy-number alterations. For every sample and chromosome, correction was carried out as follows. Because the maximum number of broad copy-number alterations on any chromosome observed in the data was two, a piecewise constant function with two pieces was fitted to the log ratios to estimate these broad alterations. Each piece was required to be longer than 5% of the chromosome length to avoid spurious small pieces. The wavy pattern was estimated by fitting a Loess curve with a window size of 100 probes to the residuals of the piecewise constant function fit. To avoid smoothing true focal alterations, the weight of probes with an absolute log ratio deviation greater than 0.5 from the piecewise constant function were set to 0 for the Loess fit. The estimated wavy pattern was then subtracted from the log ratios resulting in corrected values.

After normalization the log ratios were segmented using Circular Binary Segmentation as implemented in the R package DNACopy⁵³ (v.1.46.0). The default parameters were used except for α , which was set to 0.001. Segments with a length of 5 or less probes were merged with the closest adjacent segment. For every sample, states with means closer than 0.05 were merged iteratively beginning with the two states with closest means. When two states were merged the new mean was given by the mean of the two old state means weighted by the number of probes in every state. After state merging, the remaining segment means were adjusted to have a median of 0. These segmented copy number profiles were then deconstructed into underlying copy number events using Ziggurat Deconstruction⁵⁴. All these steps were performed using R (v.3.3.0).

Aberration events as defined by left and right change points, and aberration type were pooled across the matched samples of a single mouse to form a mouse-specific base set of aberration prototypes. For this, amplifications and deletions that were similar by more than 80% as measured by their Jaccard-index regarding probe support were merged into single prototypes using Jaccard-distance-based complete linkage clustering and union of supports. Individual primary tumour and metastases samples were then encoded according to presence (absence) of the prototypic aberrations, whereby the present prototype was called by the minimum Jaccard-distance. The resulting feature vectors were then used for phylogenetic tree inference. Phylogenetic trees were generated by assuming ideal (that is, error-free) data and inferring plausible common ancestors (intermediates) of aberration profiles by extracting shared features of an increasing number of samples, that is, evaluating common aberration events in sample pairs, triplets, quads, and so on, and organizing these ancestors according to hierarchical levels. Subsequently, admissible edges were constructed top-down between vertices allowing for two re-losses of acquired gains and no re-gains of any losses (this condition was also ensured globally for each path). Then all simple paths from the normal cell to the samples were generated using the Igraph R-package (v.1.0.0), combined into a directed acyclic graph and filtered for the fewest genomic changes along the graph and lowest number of intermediates (maximum parsimony). This resulted in one unique phylogenetic tree for each mouse.

Detection of human DCCs. For CGH analysis of single DCCs, bone-marrow sampling of patients with M1 stage cancer was performed within the study protocol of the GEBDIS study at the Central Hospital in Augsburg after informed, written consent of patients was obtained. The ethics committees of the University of Munich (ethics vote number 007/02) and Regensburg (ethics vote number 07-079) approved bone marrow sampling (including patients with M0-stage cancer) and genomic analysis of isolated cells. For all patients informed, written consent was obtained. For bone-marrow sampling and analysis for cytokeratin-positive cells of patients from Tübingen (approval by the ethics committee of Tübingen University, reference number 560/2012R) all specimens were obtained after written, informed consent. Bone-marrow sample preparation, slide preparation, cytokeratin staining and cell isolation was performed as previously described⁵.

Whole genome amplification and single cell comparative genomic hybridization. Whole genome amplification was performed as previously described^{5,55}. The method has become commercially available as a kit (*Ampli1*, Silicon Biosystems). A histogram of copy-number alterations (Fig. 5e) was generated for human primary breast cancers ($n = 1,637$) derived from the Progenetix database (<http://www.progenetix.net>) and DCCs isolated from bone marrow of breast cancer patients without (M0, $n = 94$; see Supplementary Table 7 for clinical details of patients) and with metastasis (M1, $n = 91$).

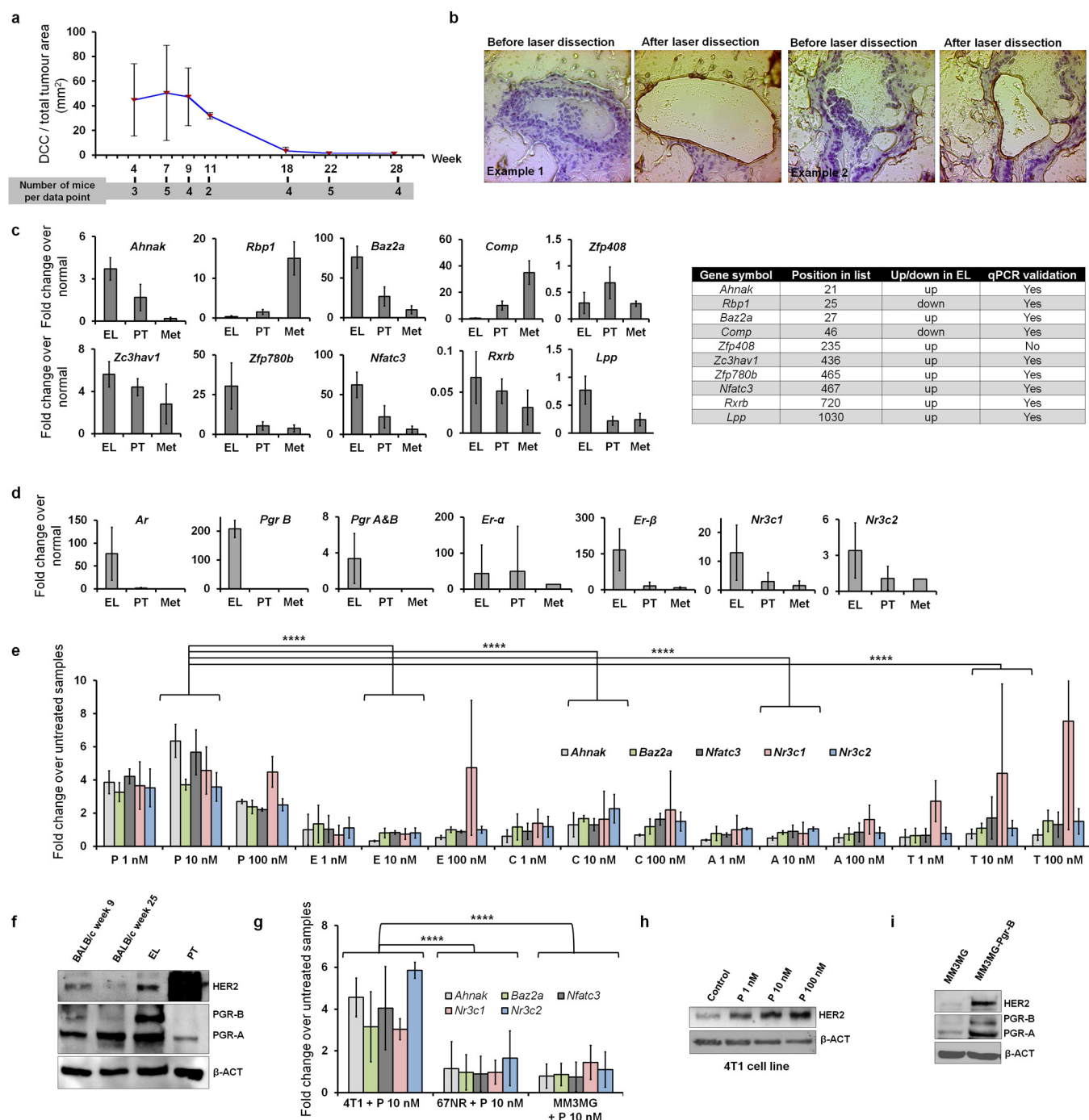
Analysis of patients with breast cancer for association of DCC and primary tumour receptor expression. We analysed data of 2,239 patients from the Department of Oncology and Obstetrics, University of Tübingen. DCC status was assessed according to the consensus protocol²³, using the anti-cytokeratin antibody A45B/B3 and by evaluating 2×10^6 bone-marrow cells. PGR expression of primary

tumours was categorized into PGR staining scores 0–1 for absent expression; 2–8 for intermediate expression and 9–12 for high expression. HER2 status of primary tumours was categorized into the staining score 0 for absence of HER2 staining (IHC negative); score 1 and score 2 without *HER2* amplification (IHC positive) and score 2 with *HER2* amplification and 3 (which is known to be caused by *HER2* amplification; FISH positive).

Bioinformatics and statistical analysis. Statistical analyses and estimation of variation within each group of data were performed using GraphPad Prism v.6 and R v.3.3.1. For *in vivo* DCC experiments of primary tumour compared to early lesion spheres, sample size was estimated using G*Power (v3). No statistical methods were used to predetermine sample size for other experiments. For each experiment, mouse numbers are given in the figures or the text. All *in vitro* and primary culture experiments were performed at least in triplicate and Student's *t*-test was used for comparisons. For all other experiments we applied the D'Agostino–Pearson omnibus normality test. When sample size was sufficiently large ($n \geq 8$) and were not distributed normally according to the D'Agostino–Pearson test ($P \leq 0.05$) we applied the Mann–Whitney *U*-test. For gene signature evaluation in Figs 1c, 3d and Extended Data Fig. 1e, g, gene wise *t*-test *P* values were combined using Stouffer's method. A linear regression test (*F*-test for slopes) was used to compare proliferation curves and tumour growth. For comparing numbers between different groups we applied Fisher's exact test or if the sample numbers were at least 5 in each condition the χ^2 test. In Fig. 1d one-way ANOVA was used. All *P* values are two-tailed. All *P* values ($0.05 \leq P \leq 0.0001$) and statistical tests are mentioned in either figures or legends. Genomatrix (v2.0) (<https://www.genomatrix.de>) was used for signalling pathway analysis and oPOSSUM (v1) (<http://opossum.cisreg.ca/oPOSSUM3/>) for transcription-factor binding-site enrichment. For miRNA-binding enrichment, we used DIANALAB (<http://diana.cslab.ece.ntua.gr/>) and for the identification of target miRNAs for single target genes the miRanda software (<http://www.microrna.org/microrna/home.do>) was used. The experiments were not randomized.

Data availability. The miRNA sequencing data and microarray data are deposited at the Gene Expression Omnibus (GEO) database under accession number GSE68683. Analysed data for microarray and miRNAs can be found in Supplementary Data 1 and 2, respectively. The mouse ancestral CGH data are deposited at the GEO database under accession number GSE87469. All raw data for presented graphs and statistics are deposited in Source Data files. Further material and data other than what is presented here can be obtained from the corresponding author (C.A.K.) upon request.

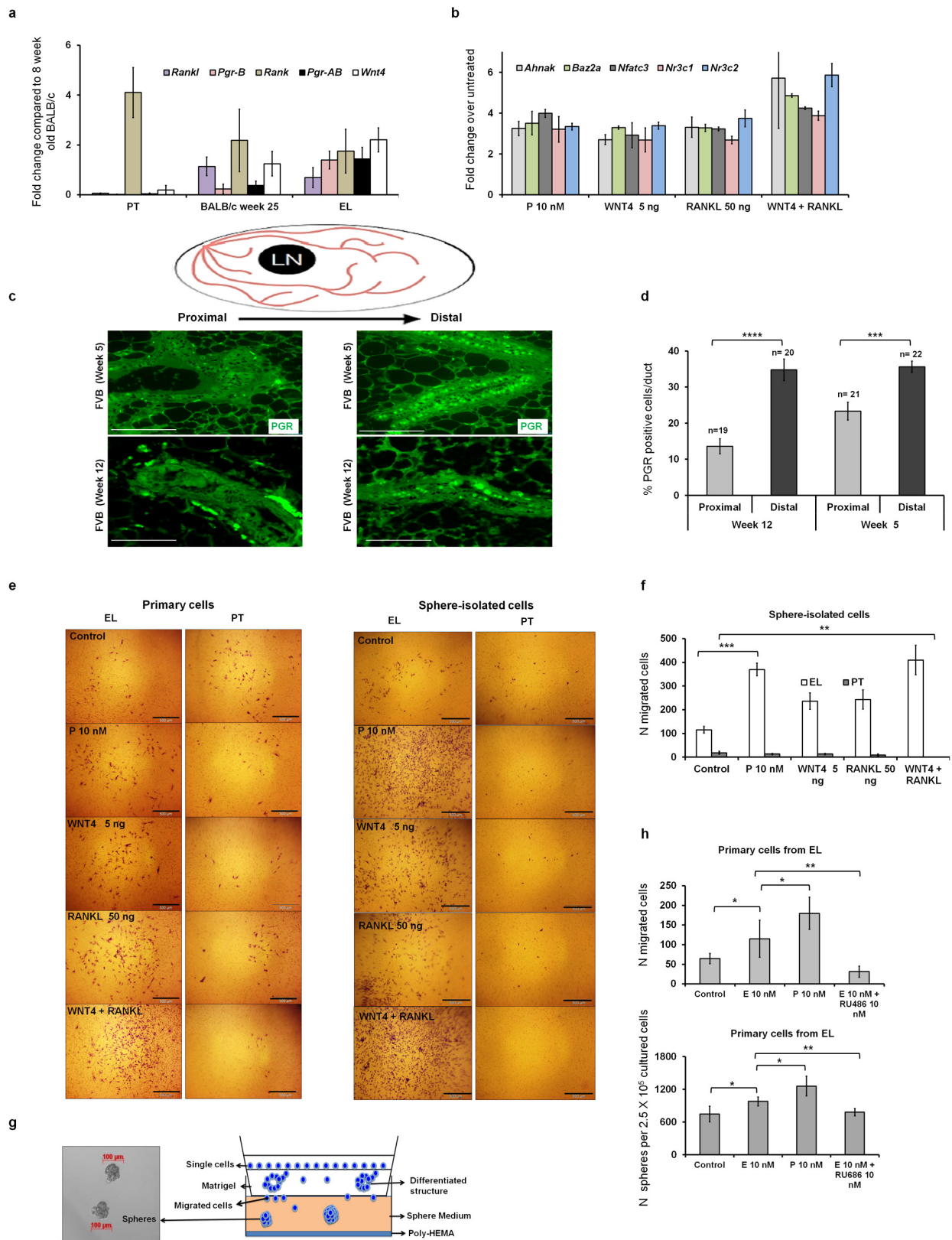
40. VerMilyea, M. D. *et al.* Transcriptome asymmetry within mouse zygotes but not between early embryonic sister blastomeres. *EMBO J.* **30**, 1841–1851 (2011).
41. Dull, T. *et al.* A third-generation lentivirus vector with a conditional packaging system. *J. Virol.* **72**, 8463–8471 (1998).
42. Petti, L. M. & Ray, F. A. Transformation of mortal human fibroblasts and activation of a growth inhibitory pathway by the bovine papillomavirus E5 oncoprotein. *Cell Growth Differ.* **11**, 395–408 (2000).
43. Stewart, S. A. *et al.* Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA* **9**, 493–501 (2003).
44. Thery, C., Amigorena, S., Raposo, G. & Clayton, A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. *Current Protoc. Cell Biology* Chapter 3, Unit 3.22 (2006).
45. Dueck, A., Eichner, A., Sixt, M. & Meister, G. A miR-155-dependent microRNA hierarchy in dendritic cell maturation and macrophage activation. *FEBS Lett.* **588**, 632–640 (2014).
46. Meyer, L. R. *et al.* The UCSC genome browser database: extensions and updates 2013. *Nucleic Acids Res.* **41**, D64–D69 (2013).
47. Yang, Y. H., Buckley, M. J. & Speed, T. P. Analysis of cDNA microarray images. *Brief. Bioinform.* **2**, 341–349 (2001).
48. Yang, Y. H. *et al.* Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30**, e15 (2002).
49. Khojasteh, M., Lam, W. L., Ward, R. K. & MacAulay, C. A stepwise framework for the normalization of array CGH data. *BMC Bioinformatics* **6**, 274 (2005).
50. Smyth, G. K. & Speed, T. Normalization of cDNA microarray data. *Methods* **31**, 265–273 (2003).
51. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
52. Marioni, J. C. *et al.* Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization. *Genome Biol.* **8**, R228 (2007).
53. Venkatraman, E. S. & Olshen, A. B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**, 657–663 (2007).
54. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
55. Klein, C. A. *et al.* Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc. Natl Acad. Sci. USA* **96**, 4494–4499 (1999).



Extended Data Figure 1 | Early lesion signature induction and expression of HER2 and PGR.

a, The proportion of cancer cells disseminating to the bone marrow in BALB-NeuT mice decreases with increasing primary tumour area. DCCs were identified using anti-cytokeratin antibodies in bone marrow samples. The y axis displays the number of detected DCCs per 10^6 bone marrow cells divided by the total tumour area in mm^2 . The number of mice used per data point is written below the graph. **b**, Laser microdissection of epithelial structures: two examples of 7–9-week-old BALB-NeuT mammary glands showing microdissection of regions with incipient epithelial hyperplasia. For all samples similar amounts of tissue (up to $100,000 \mu\text{m}^2$) were isolated. **c**, qPCR validation of microarray profiles. qPCR was performed for 10 genes, upregulated or downregulated in microarray samples of early lesions and all changes, except one (*Zfp408*), were confirmed. **d**, qPCR for the mRNA level of all steroid hormone receptors (EL, early lesions; PT, advanced primary tumour; Met, lung metastasis). **e**, Primary cultures from mammary tissue of 7–9-week-old BALB-NeuT mice, were treated

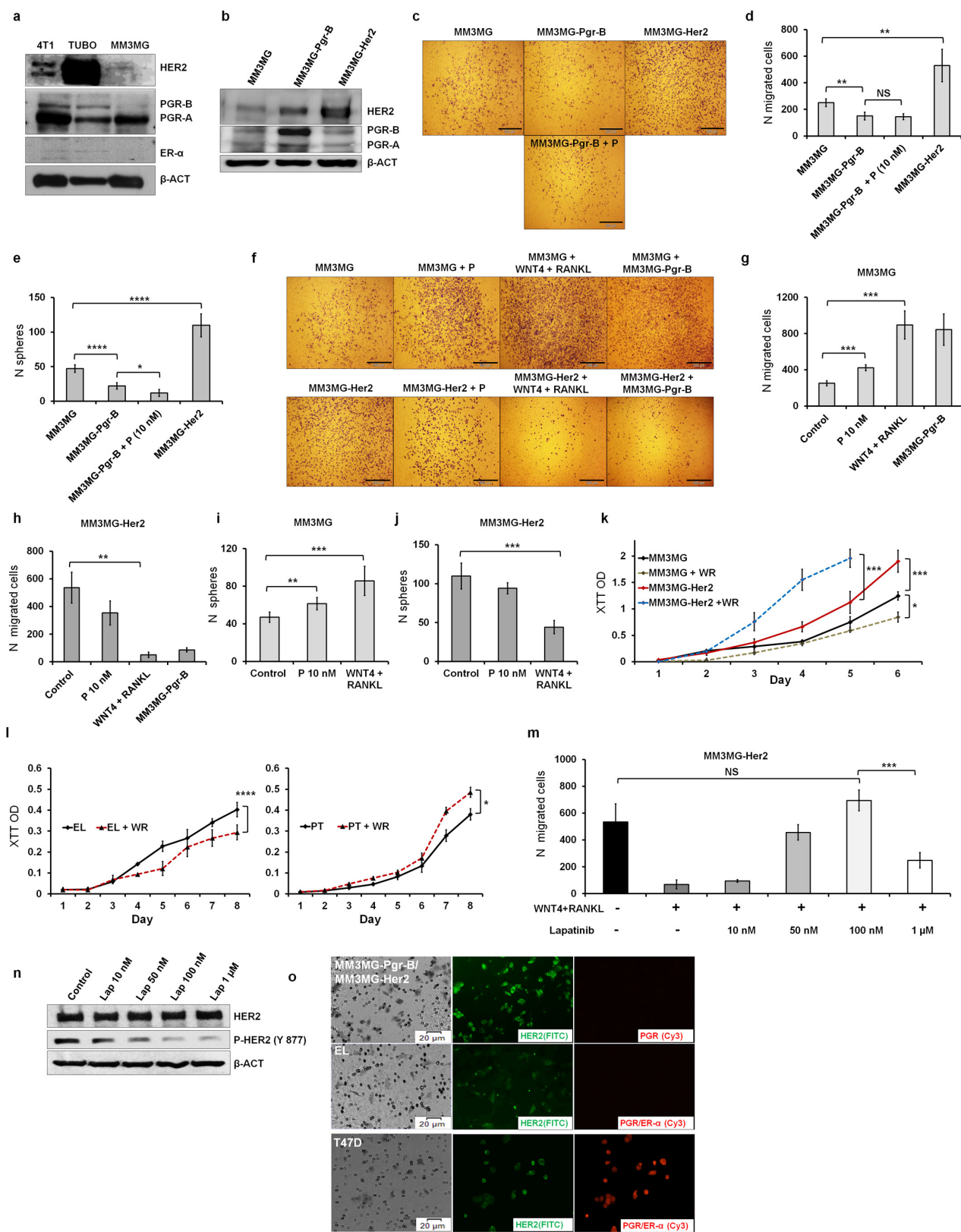
with different concentrations (1, 10, and 100 nM) of progesterone (P), oestrogen (E), aldosterone (A), cortisol (C), testosterone (T) or vehicle (ethanol; untreated) for 75 h. Only progesterone induces upregulation of the complete early lesion signature. **f**, Increased expression of PGR-B in young mammary glands (9-week old BALB/c and BALB-NeuT mice with early lesions compared to 25-week-old BALB/c mice), but not primary tumours correlates with increased HER2 expression. **g**, Progesterone induces the early lesion signature in 4T1 cells (highly aggressive and metastatic cell line derived from a spontaneous BALB/c mammary tumour), but not in 67NR (tumorigenic and non-metastatic cell line derived from a spontaneous BALB/c mammary tumour) and MM3MG cells (normal mammary epithelial cell line derived from a BALB/c mouse). **h**, Progesterone treatment induces upregulation of HER2 expression in 4T1 cells. **i**, Overexpression of PGR-B in MM3MG cells induces upregulation of HER2 expression. **** $P \leq 0.0001$ (Student's *t*-test and Stouffer's combined probability test); data are mean \pm s.d. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Progesterone regulates migration and is linked to branching morphogenesis. **a**, qPCR analysis of *Pgr*, *Rank* (also known as *Tnfrsf11a*), *Rankl* and *Wnt4* in normal and transgenic mouse mammary tissue or tumours. Note the increased expression of *Pgr*, *Wnt4* and *Rankl* in early lesions compared to primary tumours. Only *Rank* (the Rankl receptor) is strongly expressed in primary tumours. **b**, Primary cultures of early lesions treated with progesterone, WNT4, and RANKL. WNT4 and RANKL treatment induce the early lesion signature and act synergistically. **c**, **d**, PGR (green) staining at 5 and 12 weeks of age (FVB wild-type mice). Scale bars, 100 μ m. The percentage of PGR⁺ cells per duct was quantified ($n = 2$ mice per age group) in the distal and proximal portions of the gland (relative to the origin). PGR-expressing cells were more abundant in distal ducts (number of analysed ducts per group is displayed above each column). LN, lymph node. **e**, **f**, Photomicrograph of migration assay (**e**) and quantification of migrating cells (**f**) derived

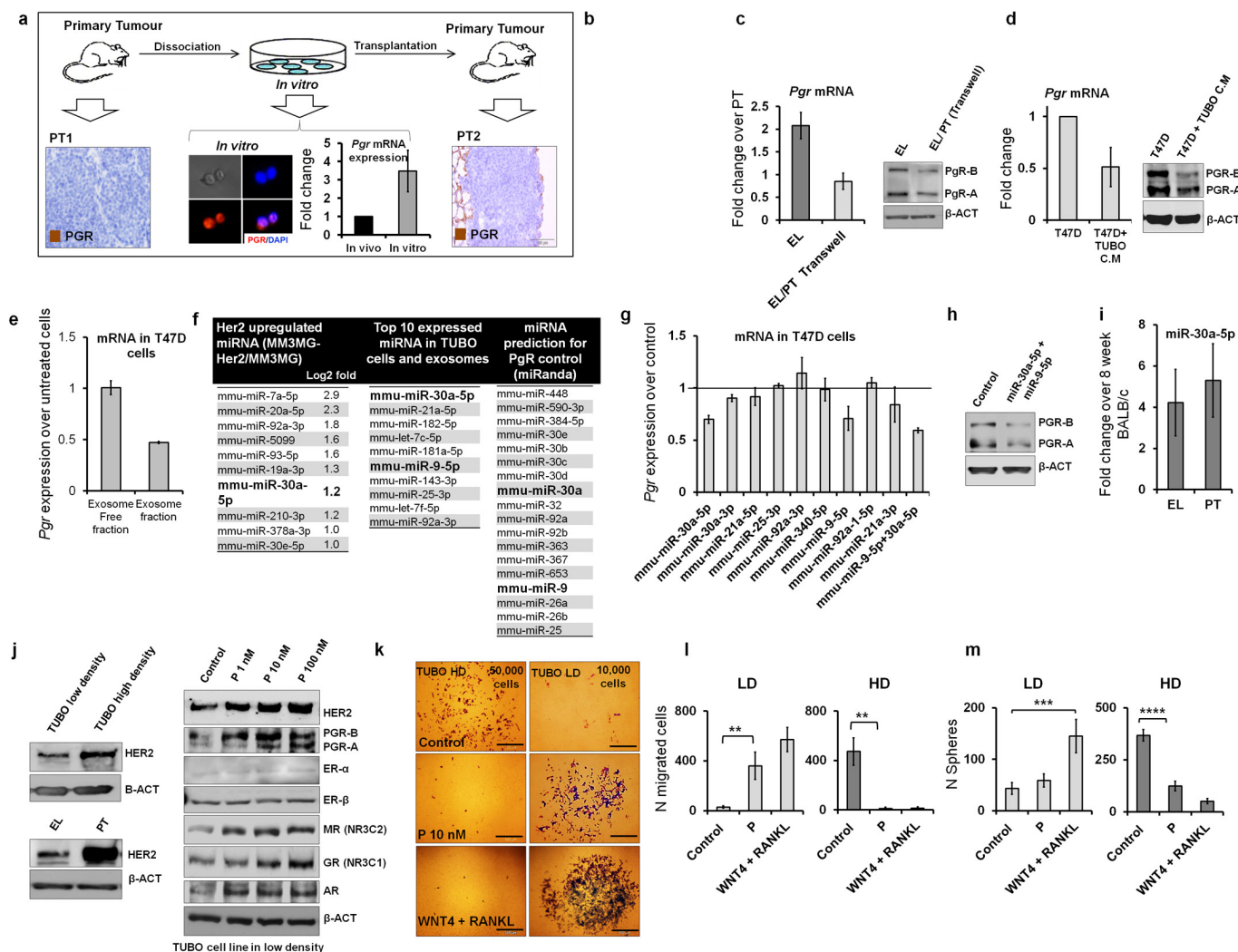
from fresh tissue (**e**, left) or dissociated spheres derived from primary tumours or early lesions (**e**, right; **f**, quantification). Progesterone, WNT4 and RANKL induce migration of early-lesion-derived but not primary-tumour-derived cells (see also Fig. 2a). **g**, Scheme of combined migration and sphere assay. The lower chamber is filled with serum-free sphere medium and the bottom is covered with poly-HEMA to prevent adhesion and enable sphere formation. After 72 h migration, the insert is removed and the lower chamber is analysed (after 11 days) for mammosphere formation (see Methods). **h**, Effect of oestrogen and progesterone on migration and sphere formation of mammary cells derived from early lesions. Cells were exposed to 10 nM oestrogen or progesterone or 10 nM oestrogen + 10 nM progesterone inhibitor (RU486). * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$ (Student's *t*-test); Data are mean \pm s.e.m. (**d**) or mean \pm s.d. (other panels).



Extended Data Figure 3 | See next page for caption.

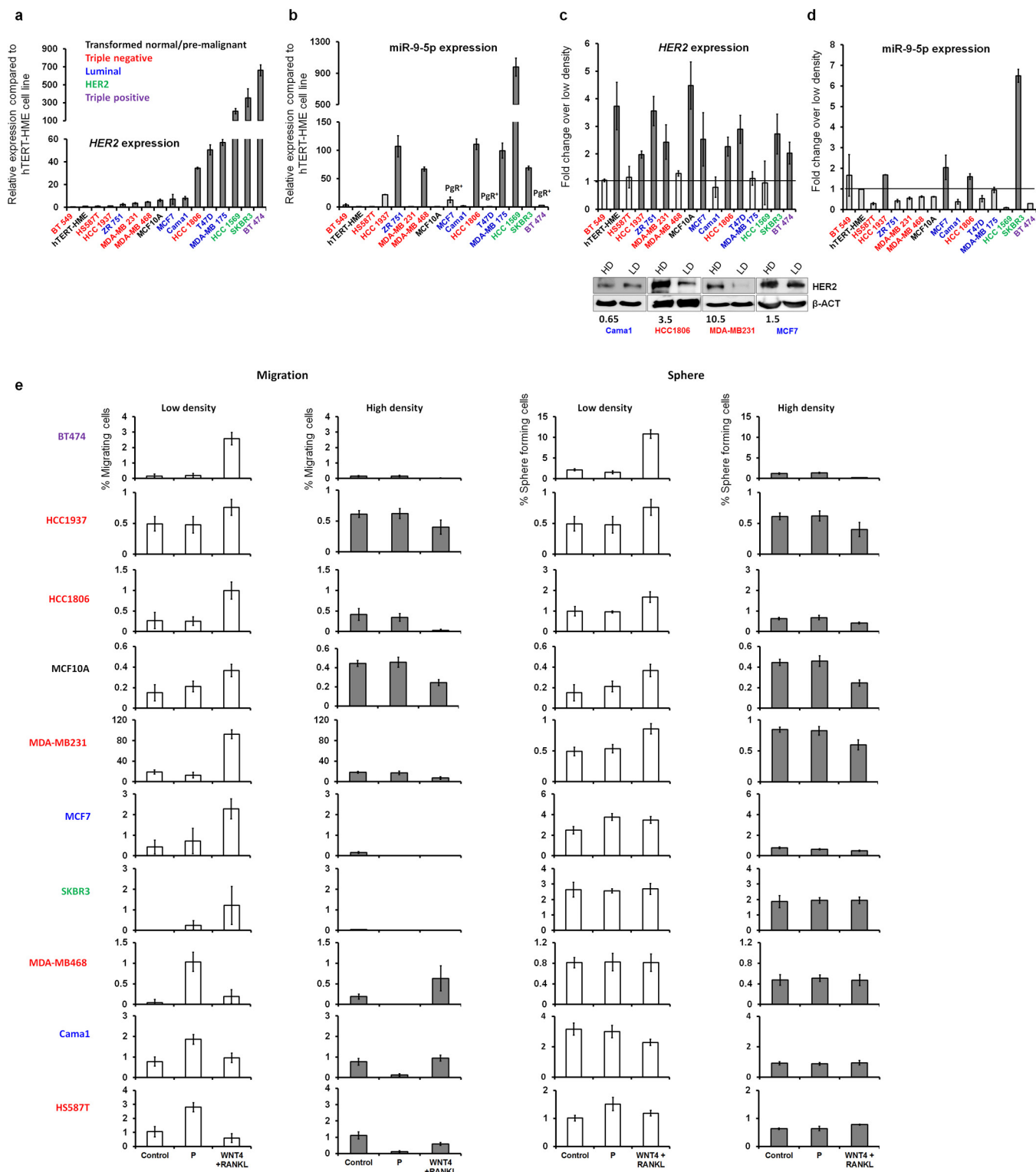
Extended Data Figure 3 | HER2 expression levels regulate migration and proliferation. **a**, Parental MM3MG cells, a cell line derived from mammary epithelial cells of wild-type BALB/c mice, do not express ER α but express low levels of HER2 and PGR-B compared to 4T1 and TUBO cells (TUBO cells were grown at low density; see Fig. 3 and Extended Data Figs 4, 5). **b**, Immunoblot confirming successful transduction of the MM3MG cell line with *Her2* and *Pgr-B*. Note that transduction of PGR-B increases HER2 levels. **c, d**, Overexpression of *Pgr-B* in MM3MG cells (MM3MG-Pgr-B) reduces migration, whereas *Her2* overexpression (MM3MG-Her2) increases migration of cells. Addition of progesterone does not alter migration of *Pgr-B*-overexpressing cells (MM3MG-Pgr-B). **e**, Overexpression of *Pgr-B* in MM3MG cells reduces sphere formation, whereas *Her2* overexpression increases sphere formation. **c–e**, Migrating/sphere forming cells are not from the PGR⁺, but the PGR[−] population, which are responsive to PIPS. **f–j**, To investigate which PGR[−] cells were the target population of progesterone signalling, we exposed parental MM3MG cells and *Her2*-transduced cells to progesterone, PIPS or mixed the cells with PGR⁺ cells (only for migration experiments). Progesterone, WNT4 and RANKL, and co-culture with MM3MG-Pgr-B induced sphere formation and migration of MM3MG cells, but decreased these responses in MM3MG-Her2 cells. **k**, Overexpression of *Her2* increases proliferation of MM3MG cells (MM3MG-Her2). WNT4 and RANKL (WR) further

increase proliferation of MM3MG-Her2 cells, but decreases proliferation of the parental (MM3MG) cells. Therefore based on expression of HER2, cells either migrate (HER2^{low/−}) or proliferate (HER2^{high}). **l**, WNT4 and RANKL treatment induces proliferation of primary cultured cells derived from primary tumours, but reduces it in cells derived from early lesions. **m**, Reduction of HER2 signalling by lapatinib overrides the inhibitory effect of WNT4 and RANKL, and increases migration in MM3MG-Her2 cells. However, strong inhibition of HER2 signalling reduces migration. **n**, Lapatinib inhibits HER2 signalling by preventing phosphorylation. **o**, Cells that migrated through the pores of the migration chamber insert were stained for HER2 (FITC, green) and PGR or ER α (Cy3, red). In 1:1 co-culture of MM3MG-Pgr-B and MM3MG-Her2 (top) only HER2-expressing cells migrate. Migrated primary cells derived from early lesions (middle) do not express PGR and display faint HER2 staining (brightness of HER2 and PGR staining increased by 50% for better visibility). HER2 and PGR double-positive T47D cells fixed onto the filters of migration chambers served as positive control of staining. **m–o**, Cells with low/intermediate signalling of HER2 show the highest response in migration and sphere formation induction by PIPS. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$; NS, not significant; (*F*-test of the slope (**k, l**) or Student's *t*-test (other panels)); data are mean \pm s.d. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 4 | Cell density and regulation of PGR and HER2 signalling. **a**, PGR expression silenced in tumours can be re-activated in culture and re-silenced *in vivo*. **b**, PGR re-expression only occurs in TUBO cells grown at low density and high density after frequent medium change. **c**, Downregulation of PGR in early lesion cells cultured in a transwell assay next to primary tumour cells suggests the existence of a secreted factor passing through the membrane of the transwell insert and downregulating *Pgr* mRNA (left) and protein (right). **d**, T47D cells exposed to conditioned medium from TUBO cells display reduced *Pgr* mRNA (left) and protein (right). **e**, Exosomes derived from the cell culture medium of TUBO cells grown at high density (exosome fraction) induce downregulation of PGR in T47D cells. **f**, miRNA sequencing to identify PGR regulating miRNAs. Left, top 10 upregulated miRNAs in *Her2*-overexpressing cells (MM3MG-Her2) compared to control (MM3MG). Middle, top 10 expressed miRNAs in TUBO cells and TUBO-cell-derived exosomes. Right, miRNAs predicted by the miRanda web software to regulate *Pgr* **g**, Among all

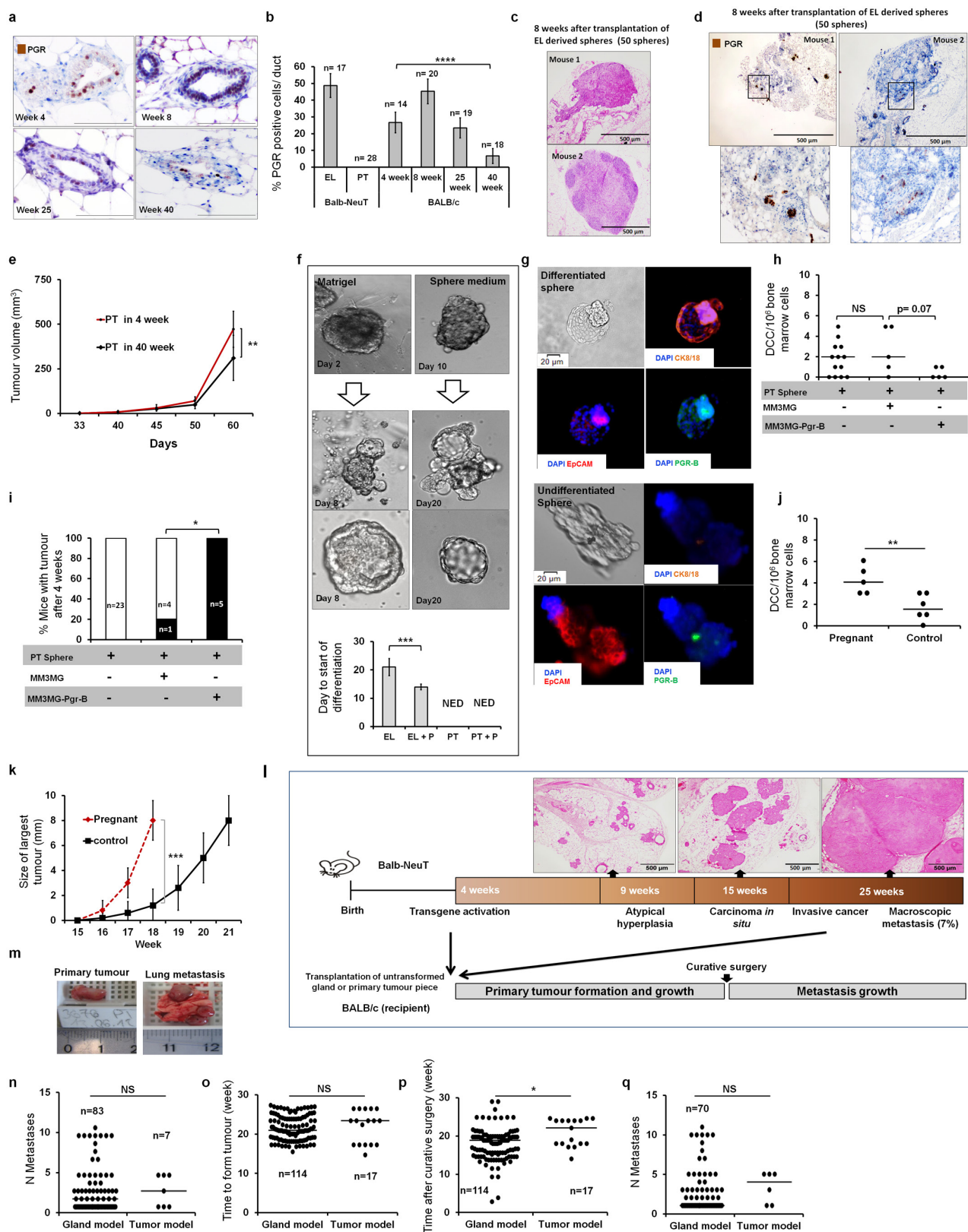
candidate miRNAs only miR-30a-5p and miR-9-5p induce downregulation of *Pgr* mRNA in T47D cells. **h**, Downregulation of PGR in T47D cells treated with miR-30a-5p and miR-9-5p. **i**, Expression of miR-30-5p in early lesions and primary tumour samples compared to mammary glands from 8-week-old BALB/c mice. **j**, Density induced upregulation of HER2 in TUBO cells grown at low or high density and early lesions compared to primary tumour samples (left) and progesterone-responsiveness of low-density TUBO cells (right). Note that levels of HER2, PGR, the glucocorticoid (GR), mineralocorticoid (MR) and androgen receptors (AR) are regulated by progesterone in a dose-dependent manner except ER- α (see related result in Extended Data Fig 1d). **k–m**, TUBO cells grown at low density and exposed to progesterone or PIPS migrated more (**k**, **l**) and produced more spheres (**m**), similar to cells derived from early lesions (see Fig. 2). $^{**}P \leq 0.01$; $^{***}P \leq 0.001$; $^{****}P \leq 0.0001$ (Student's *t*-test); data are mean \pm s.d. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Cell density and regulation of PGR and HER2 signalling in human cell lines. To investigate whether human breast cancer cells display similar regulatory circuits as found in mouse cells, we selected 16 cell lines of different breast cancer subtypes. **a**, *HER2* mRNA expression levels in 15 human cell lines compared to the hTERT-HME cell line. Different colours indicate the subtype of breast cancer of each cell line. **b**, The expression of miR-9-5p in human breast cancer cell lines compared to hTERT-HME cells. Note that cell lines highly expressing *HER2* (see **a**) express more miR-9-5p similar to primary tumours of BALB-NeuT and TUBO cells (see Fig. 3b, c), whereas two *HER2*^{high}/PGR^{high} cell lines, BT474 and T47D, do not express miR-9-5p similar to human *HER2*^{high}/PGR^{high} samples (see Fig. 5d). **c**, High cell density leads to upregulation of *HER2* mRNA (top) or protein levels (bottom) in several cell lines. Only four cell lines were analysed for protein level (*HER2* level was not influenced by cell density in CAMA1; *HER2* level regulated by cell density in HCC1806, MDA-MB-231 and MCF7). Numbers below the blots denote fold change of *HER2* in high density compared to low density

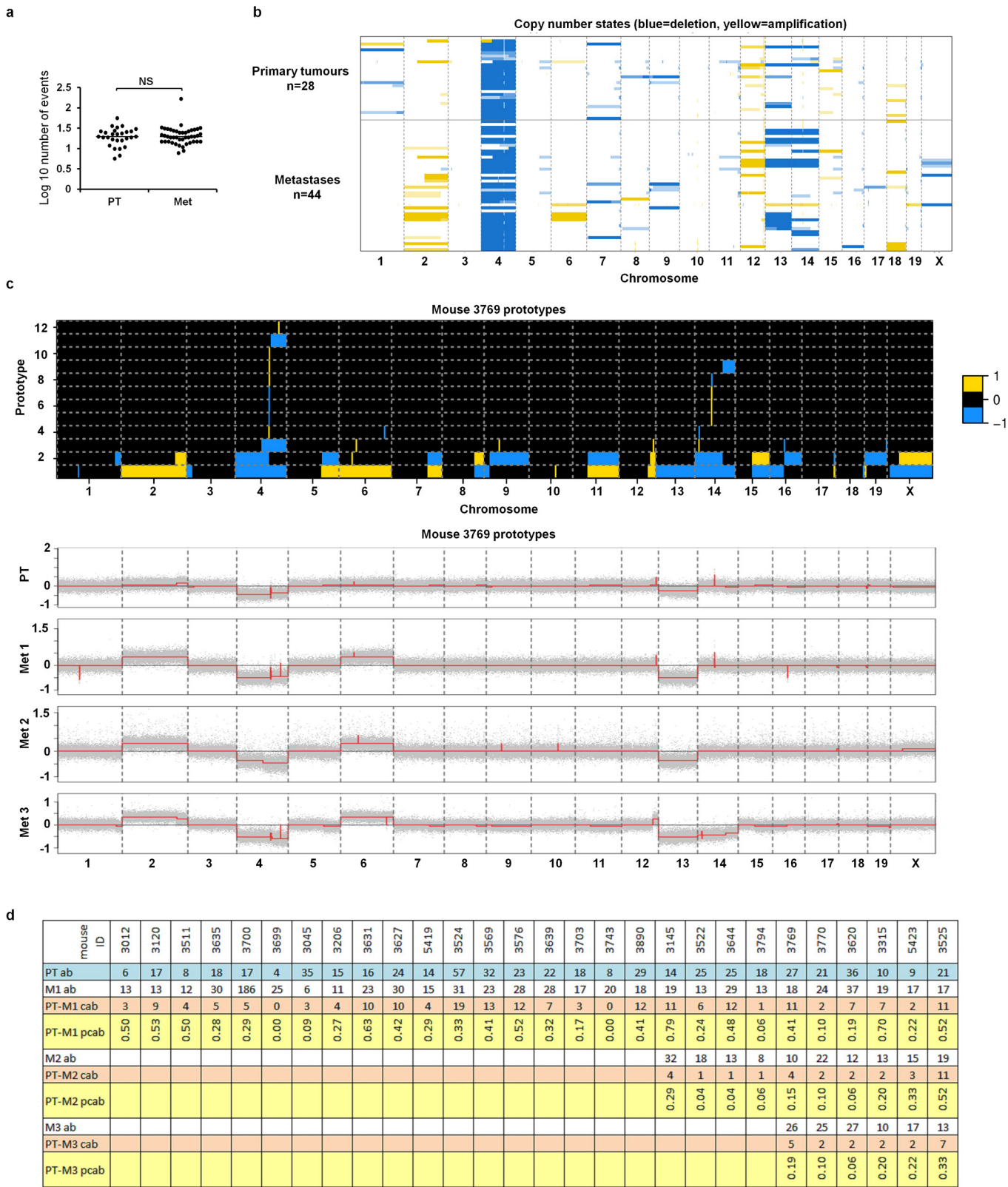
normalized over β -actin. **d**, Expression of miR-9-5p is upregulated by cell density in SKBR3, HCC1937, HCC1806 and MCF7 cell lines. **e**, Migration and sphere-forming potential of 10 out of 16 cell lines grown at low and high densities, and treated with PIPS or progesterone. The first 7 cell lines regulate *HER2* transcripts by density (see **c**) and their response to PIPS is similar to TUBO cells and primary mammary cell cultures of BALB-NeuT mice (see Fig. 2 and Extended Data Fig. 4l, m). The remaining three cell lines do not regulate *HER2* transcripts by cell density, but respond to progesterone similarly to the TUBO cell line and primary mammary cell cultures of BALB-NeuT mice (see Fig. 2 and Extended Data Fig. 4l, m). We did not perform functional assays with BT549 (triple negative subtype), T47D (luminal, MCF7-like), MDA-MB-175, ZR75-1 (luminal, CAMA1-like), hTERT-HME (transformed normal, similar to MCF10A) because of breast cancer subtype redundancy or poor growth (HCC1569). *y* axes show the percentage of migrating cells (left) and observed spheres (right) relative to seeded cells. Data are mean \pm s.d. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | Differentiation ability and metastasis formation. **a, b**, Representative images (**a**) and quantification (**b**) of PGR expression of mammary epithelial cells from wild-type BALB/c mice at 4, 8, 25 and 40 weeks of age. Scale bars, 100 μ m. PGR expression was reduced by 75% in 40-week-old wild-type mammary gland compared to 4-week-old mice and disappeared in primary tumours (see also Extended Data Figs 1f, 2a). **n**, number of ducts or glands (in early lesions and normal tissue) or visual fields in primary tumours. **c, d**, Representative micrographs of lesions 8 weeks after transplantation of early lesion spheres resembling DCIS (**c**) or less-advanced early lesions (**d**) displaying PGR expression (brown nuclear staining). **e**, Tumour growth from primary tumourspheres in young and old recipients. **f**, Differentiation of cells from early lesions but not of primary tumour cells in Matrigel (left) or in sphere culture (right) into acinus-like structures. Progesterone stimulation accelerated formation of acinus-like structures by early lesion cells, under mammosphere conditions. NED, no evidence of differentiation. **g**, Staining for CK8/18, PGR and EpCAM (epithelial cell adhesion molecule) shows expression of PGR and CK8/18 only in differentiated structures (top) compared to undifferentiated spheres (bottom). **h, i**, Primary tumourspheres were transplanted alone ($n = 23$) or co-transplanted with MM3MG-Pgr-B ($n = 5$) or MM3MG spheres ($n = 5$). DCCs numbers in bone marrow (**h**) and the number of mice with tumours (**i**) were checked 4 weeks later. Pgr-B-transduced mammary epithelia suppressed metastatic dissemination and accelerated tumour formation from primary tumourspheres. **j**, Pregnancy at the early lesion

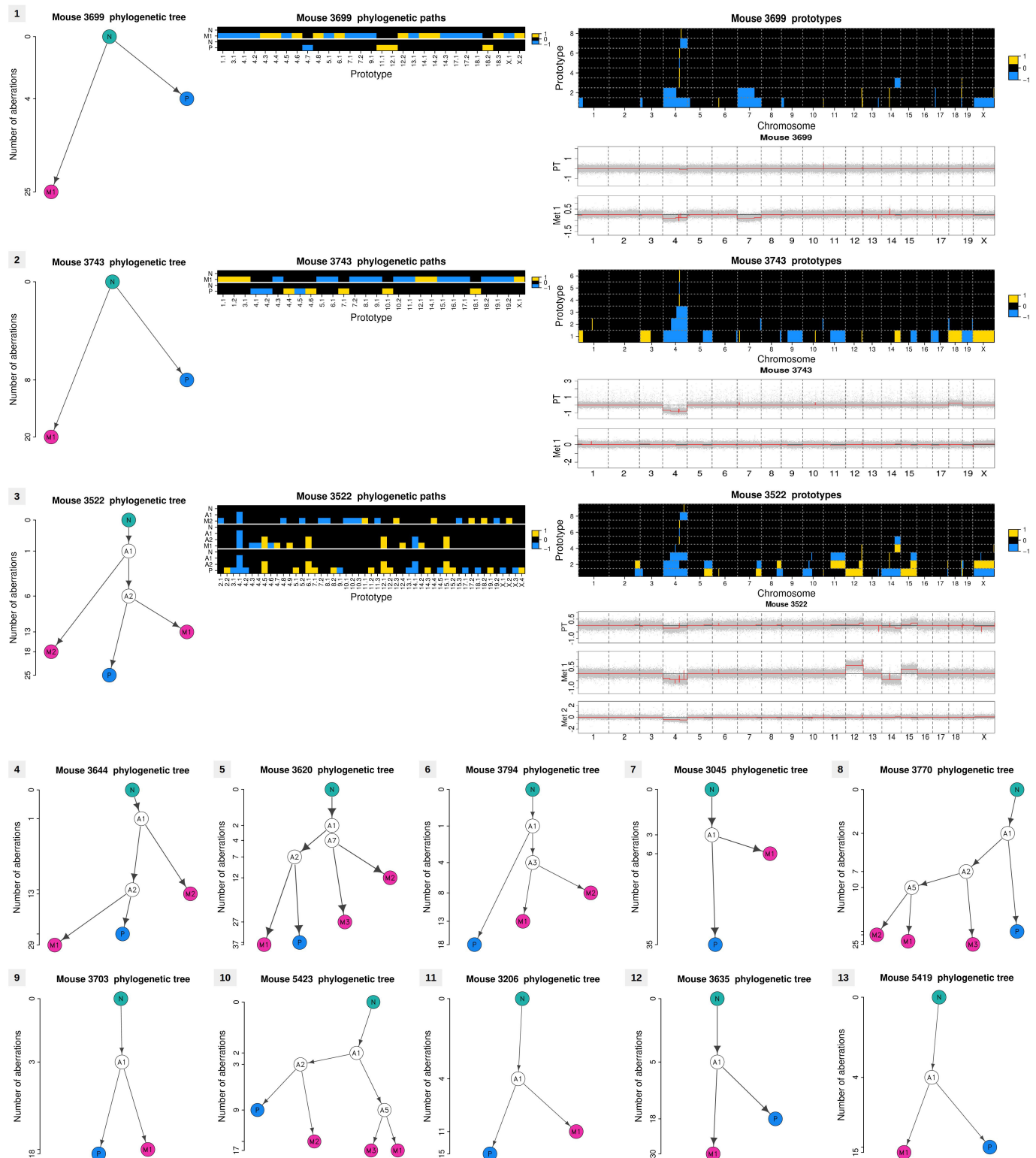
stage induces dissemination. A group of young BALB-NeuT mice mated ($n = 5$) at the early lesion stage (7-week-old) and were killed at the end of pregnancy. These mice did not form palpable tumours, but had a higher number of DCCs compared to unmated control mice ($n = 6$). **k**, Pregnancy at the advanced tumour stage. A group of BALB-NeuT mice ($n = 5$) were mated at the time of *in situ* carcinoma (15-week-old) and killed at the end of pregnancy. All pregnant mice had faster growing tumours compared to unmated control mice. **l**, Schematic of transplantation protocol for mammary gland or primary tumour tissue pieces into wild-type recipients. **m**, Example of primary tumour and macro-metastasis assessment. **n**, Number of metastatic foci in transplanted mice. 18 mice from the gland model and 3 mice from the tumour model were excluded from analyses owing to the fusion of metastatic lesions, which made it difficult to count individual lesions. **o**, Similar growth kinetics of primary tumours from gland and tumour piece models for samples from the red box in Fig. 4e. **p, q**, Mice from **o** were compared for the duration of the follow-up period after surgery. Mice from both groups were killed at the first signs of general health deterioration, which occurred earlier in gland-model mice (**p**). Longer follow-up time after curative surgery did not result in more metastases in recipients transplanted with primary tumour pieces (**q**). * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$; NS, not significant; (Student's *t*-test (**b, f, h, j**); Fisher's exact test (**i**); *F*-test for the slopes (**c, k**); Mann-Whitney *U*-test (**n-q**)). Data are shown as mean \pm s.d. (**b, e, f, k**) or median (**h, j, n-q**).



Extended Data Figure 7 | See next page for caption.

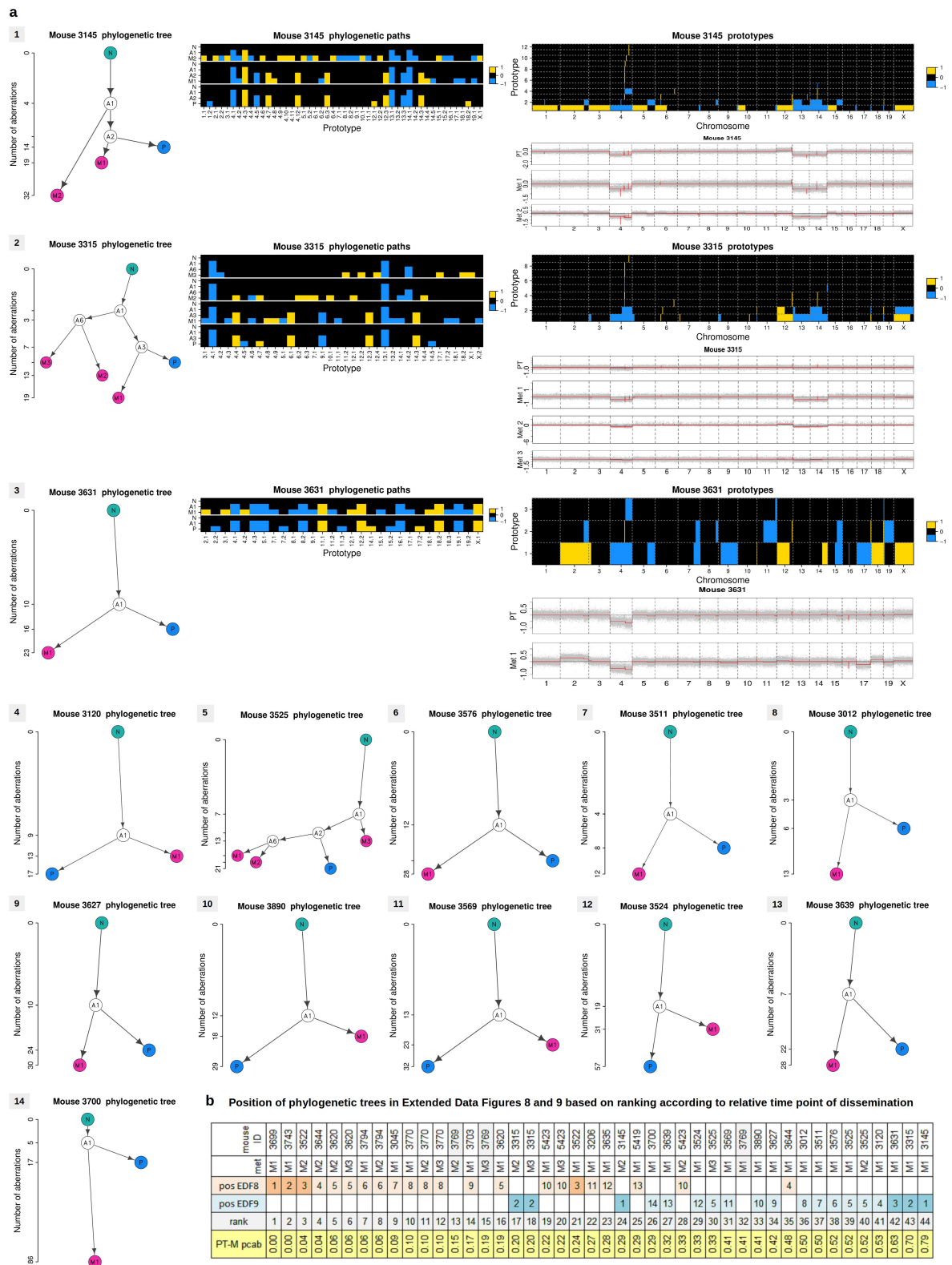
Extended Data Figure 7 | Array CGH analysis of primary tumour–metastasis pairs. **a**, Number of aberrations detected by array comparative genomic hybridization in primary tumours and matched lung metastases. Dot plot with median; statistical analysis by Mann–Whitney *U*-test. **b**, Heatmap of copy number states for the 28 primary tumours and 44 matched metastases across chromosomes 1–19 and X. Light, medium and dark yellow or blue colours indicate weak, intermediate and strong amplification (yellow) or deletion (blue) amplitudes, respectively (thresholds at ± 0.1 , ± 0.2 , ± 0.3). **c**, Prototype aberrations (top) constructed from segmented array CGH profiles (bottom) of the primary tumour (PT) and the matched metastases (Met 1–3) of mouse 3769 (phylogenetic tree and phylogenetic paths displayed in Fig. 4h, i). Prototypes (top) are organized in stacked rows per chromosome and numbered according to chromosome and positional order of their

first change point, for example, 1.2 denotes the second prototype of chromosome 1. These prototype aberrations are then used to construct the phylogenetic paths (for example, Fig. 4i) and trees. For better visibility, small focal aberrations were enlarged to have a minimal extension of 300 probes. Yellow, amplification (+1); blue, deletion (−1). Corresponding segmentation profiles of the normalized and wavy-pattern-corrected array CGH data (grey dots) are indicated by red lines (bottom). For segmentation and prototype construction, see Methods. **d**, Table for calculating the relative time points of dissemination (Fig. 4j). PT ab, number of aberrations in the primary tumour; M_k ab, number of aberrations in the matched metastases ($k = 1, 2, 3$); PT– M_k cab, number of common aberrations between primary tumours and metastases; PT– M_k pcab, proportion of common aberrations relative to the primary tumour, that is, $pcab = cab/PT\ ab$.



Extended Data Figure 8 | Phylogenetic analysis of metastasis (early divergence). Phylogenetic trees of the top 13 out of 28 primary tumours and matched metastasis samples listed according to earliest time point of dissemination (for details see tables in Extended Data Figs 7, 9). A1–7, inferred common ancestors (intermediates); M, metastases (M1–3); N, normal cell; P, primary tumour. (see Methods). The ordinate indicates the number of aberrations per profile (on a square root scale). For the first three matched samples, the phylogenetic tree paths

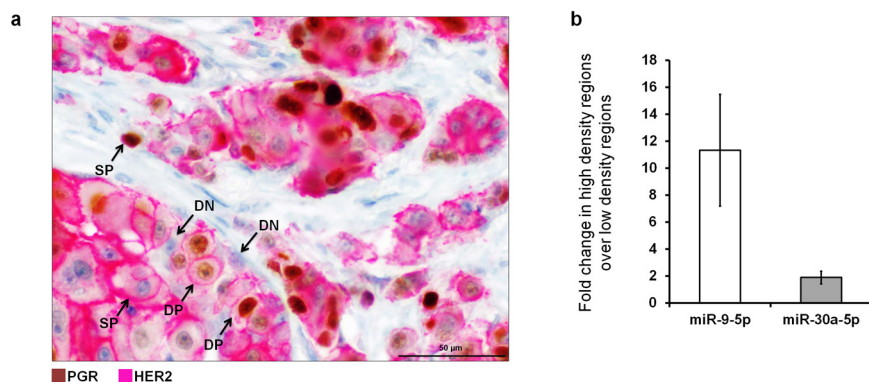
(middle), prototype aberrations (top right) and segmented array CGH profiles (bottom right) are also shown in addition to the phylogenetic tree (left). Aberration profiles along phylogenetic paths run from N via A1–7 to P or M1–3. Aberration prototypes are named according to chromosome and positional order of their first change point, for example, 2.2 denotes the second aberration prototype of chromosome 2 (see Methods and Extended Data Fig. 7c).



Extended Data Figure 9 | Phylogenetic analysis of metastasis

(late divergence). a, Phylogenetic trees of the top 14 out of 28 primary tumour and matched metastasis samples listed according to latest time point of dissemination. For the first three mice the phylogenetic paths (mid), prototype aberrations (top right) and segmented array CGH profiles (bottom right) are also shown next to the phylogenetic tree (left). See Methods and Extended Data Figs 7, 8. **b**, Summary table of all phylogenetic analyses indicates the position of the corresponding mouse phylogenetic tree in Extended Data Figs 8 and 9 (EDF8 and EDF9) for each primary tumour–metastasis pair. The two bottom rows indicate the rank

and the corresponding relative time point of dissemination as measured by the proportion of aberrations shared between primary tumour and metastasis (PT–M pcab; see Extended Data Fig. 7 and Fig. 4j). Note that only metastases ranked on position 36–44 diverged late. The phylogenetic tree and phylogenetic paths for mouse 3769 are displayed in Fig. 4h, i. In the pos-EDF8 and pos-EDF9 rows the darker colours are samples of which all data, including phylogenetic paths, prototype aberrations, segmented array CGH profiles and phylogenetic trees are shown. Faint colour cells in pos-EDF8 and pos-EDF9 are samples where only phylogenetic trees are shown.



Extended Data Figure 10 | PGR, HER2 signalling and dissemination in breast cancer patients. **a**, Double staining of a $HER2^{high}/PGR^{high}$ human breast cancer sample (PGR, brown, nucleus; HER2, red/pink, membrane). Cells with varying expression levels of HER2 and PGR, as well as negative, single- or double-positive cells can be seen. Scale bar, 100 μ m. Arrows

indicate, double-positive (DP), double-negative (DN) and single-positive (SP) cells. **b**, Lack of PGR expression in high-density areas of $HER2^{high}/PGR^{high}$ -classified tumour samples (see Fig. 5c) is directly linked to high miR-9-5p and miR-30a-5p expression. Data are mean \pm s.d.

Evidence for a spinon Fermi surface in a triangular-lattice quantum-spin-liquid candidate

Yao Shen¹, Yao-Dong Li², Hongliang Wo¹, Yuesheng Li³, Shoudong Shen¹, Bingying Pan¹, Qisi Wang¹, H. C. Walker⁴, P. Steffens⁵, M. Boehm⁵, Yiqing Hao¹, D. L. Quintero-Castro⁶, L. W. Harriger⁷, M. D. Frontzek⁸, Lijie Hao⁹, Siqin Meng⁹, Qingming Zhang^{3,10,11}, Gang Chen^{1,11,12} & Jun Zhao^{1,11}

A quantum spin liquid is an exotic quantum state of matter in which spins are highly entangled and remain disordered down to zero temperature. Such a state of matter is potentially relevant to high-temperature superconductivity and quantum-information applications, and experimental identification of a quantum spin liquid state is of fundamental importance for our understanding of quantum matter. Theoretical studies have proposed various quantum-spin-liquid ground states^{1–4}, most of which are characterized by exotic spin excitations with fractional quantum numbers (termed ‘spinons’). Here we report neutron scattering measurements of the triangular-lattice antiferromagnet YbMgGaO₄ that reveal broad spin excitations covering a wide region of the Brillouin zone. The observed diffusive spin excitation persists at the lowest measured energy and shows a clear upper excitation edge, consistent with the particle–hole excitation of a spinon Fermi surface. Our results therefore point to the existence of a quantum spin liquid state with a spinon Fermi surface in YbMgGaO₄, which has a perfect spin-1/2 triangular lattice as in the original proposal⁴ of quantum spin liquids.

In 1973, Anderson⁴ proposed the idea of a quantum spin liquid (QSL) in the study of the triangular-lattice Heisenberg antiferromagnet. This idea was revived after the discovery in 1986 of high-temperature superconductivity⁵. A QSL, as currently understood, does not fit into Landau’s conventional paradigm of symmetry-breaking phases^{1,2,6,7}, and is instead an exotic state of matter characterized by spinon excitations and emergent gauge structures^{1–3,6}. The search for QSLs in models and materials^{8–12} has been partly facilitated by the Oshikawa–Hastings–Lieb–Schultz–Mattis (OHLMS) theorem, which hints at the possibility of QSLs in Mott insulators with odd electron fillings and a global U(1) spin rotational symmetry^{13–15}. Indeed, a continuum of spin excitations has been observed in the kagome-lattice material ZnCu₃(OD)₆Cl₂ (refs 12, 16). However, the requirement of the U(1) spin rotational symmetry prevents the application of the OHLMS theorem in strongly spin–orbit-coupled Mott insulators, in which the spin rotational symmetry is completely absent. A recent theory addressed this limitation of the OHLMS theorem, arguing that, as long as time-reversal symmetry is preserved, the ground state of a spin–orbit-coupled Mott insulator with odd electron fillings must be exotic¹⁷.

The triangular antiferromagnet YbMgGaO₄ (refs 18, 19) displays no indication of magnetic ordering or symmetry breaking at temperatures as low as 30 mK, despite the energy scale for spin interaction being equivalent to a temperature of approximately 4 K. Because of the strong spin–orbit coupling of the Yb electrons, YbMgGaO₄ was the first QSL to be proposed that was unlike those in the OHLMS theorem¹⁹. The

thirteen 4f electrons of the Yb³⁺ ion form the spin–orbit-entangled Kramers doublets that are split by the D_{3d} crystal electric fields^{20–22}. At temperatures considerably lower than the crystal field gap (about 420 K), the magnetic properties of YbMgGaO₄ are captured by the ground-state doublet that is described by an effective spin-1/2 local moment; this is confirmed by a measured magnetic entropy of $R\ln(2)$ per Yb³⁺ ion¹⁸, where R is the ideal gas constant. Figure 1a, b shows that the YbO₆ octahedra form well-separated triangular layers. Because of the large difference in chemistry between Yb³⁺ and the non-magnetic Mg²⁺/Ga³⁺ ions, intra-triangular-layer impurities are prevented in YbMgGaO₄ (refs 18, 19, 21). Hence, the Yb system is a spin-1/2 antiferromagnet on a perfect triangular lattice.

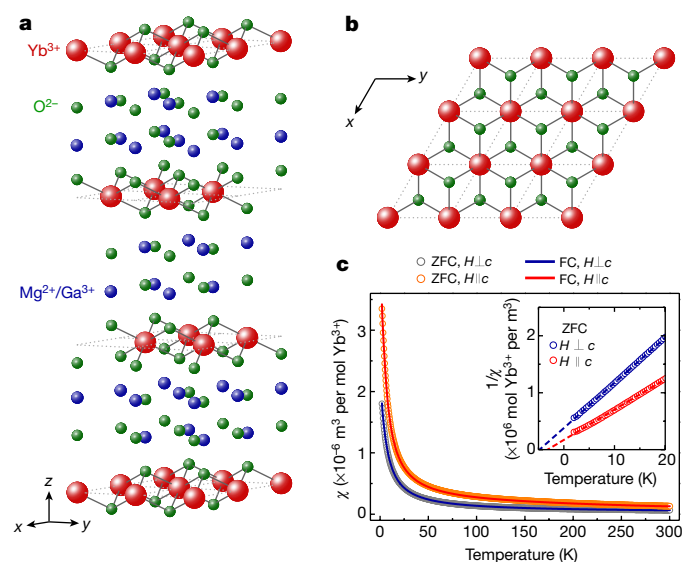


Figure 1 | Crystal structure and magnetic susceptibility of a single crystal of YbMgGaO₄. **a**, Schematic of the YbMgGaO₄ crystal structure. The dashed line indicates the unit cell. **b**, A triangular layer of Yb³⁺ ions and oxygen. **c**, Direct-current magnetic susceptibility χ measured under zero-field cooling (ZFC) and field cooling (FC) for single crystals of YbMgGaO₄, under magnetic fields ($H = 1$ T) applied perpendicular and parallel to the c axis. Paramagnetic behaviour is observed at low temperature with no obvious differences between ZFC and FC data. The inset shows the inverse susceptibility $1/\chi$ at low temperature (≤ 20 K), fitted with the Curie–Weiss law (dashed line). The fitting results in Curie temperatures of $\theta_{CW,\perp} = -4.78$ K and $\theta_{CW,\parallel} = -3.2$ K for perpendicular and parallel magnetic fields, respectively.

¹State Key Laboratory of Surface Physics and Department of Physics, Fudan University, Shanghai 200433, China. ²School of Computer Science, Fudan University, Shanghai 200433, China.

³Department of Physics, Beijing Key Laboratory of Opto-electronic Functional Materials and Micro-nano Devices, Renmin University of China, Beijing 100872, China. ⁴ISIS Facility, Rutherford

Appleton Laboratory, STFC, Chilton, Didcot OX11 0QX, UK. ⁵Institut Laue-Langevin, 71 Avenue des Martyrs, 38042 Grenoble Cedex 9, France. ⁶Helmholtz-Zentrum Berlin für Materialien und

Energie, D-14109 Berlin, Germany. ⁷NIST Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA. ⁸Quantum Condensed Matter

Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6393, USA. ⁹Neutron Scattering Laboratory, China Institute of Atomic Energy, Beijing 102413, China. ¹⁰Department of

Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China. ¹¹Collaborative Innovation Center of Advanced Microstructures, Nanjing 210093, China. ¹²Center for Field Theory and Particle Physics, Fudan University, Shanghai 200433, China.

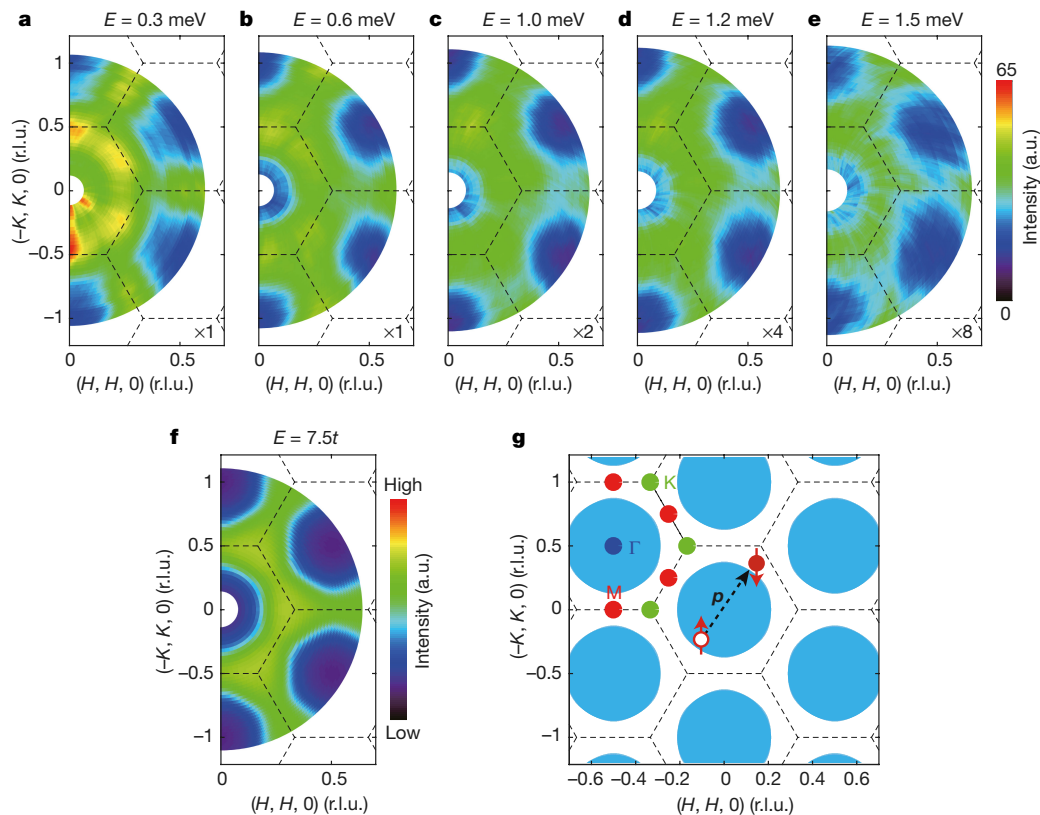


Figure 2 | Measured and calculated momentum dependence of the spin excitations, and calculated spinon Fermi surface of YbMgGaO₄.

a–e, Constant-energy images at the indicated energies and $T = 70$ mK, displaying diffusive magnetic excitations covering a wide region of the Brillouin zone. The scattering intensity is represented by the colour scale, and in **c**, **d** and **e** has been multiplied by 2, 4 and 8, respectively, for clarity. The data were collected on ThALES using the Flatcone detector, and were corrected for neutron-beam self-attenuation (Methods). **f**, Calculated momentum dependence of the spin excitations for a typical, finite E .

To characterize the behaviour of the local moment of Yb, we first measured the magnetic susceptibility of single-crystalline YbMgGaO₄ (Fig. 1c). For magnetic fields H applied both parallel to and normal to the c axis of the lattice, we found predominantly antiferromagnetic spin interactions, as evidenced by negative Curie–Weiss temperatures (Fig. 1c, inset). Because of the anisotropic spin interaction, the Curie–Weiss temperatures for $H \perp c$ and $H \parallel c$ were not identical (Fig. 1c, inset; Extended Data Fig. 1f), with $\Theta_{\text{CW},\perp} = -4.78$ K and $\Theta_{\text{CW},\parallel} = -3.2$ K, consistent with previous measurements^{18,19}. We examined the magnetic susceptibilities in field cooling (FC) and zero-field cooling (ZFC) measurements. No splitting was detected between the FC and ZFC results down to 2 K, indicating the absence of spin glassy transitions (Fig. 1c).

The Curie–Weiss temperature and the spin excitation bandwidth (discussed below) set the energy scale for the spin interactions. Our elastic neutron scattering measurements revealed no magnetic Bragg peaks (Extended Data Fig. 2) at temperatures as low as 30 mK, considerably lower than the Curie–Weiss temperature (about 4 K) and spin excitation bandwidth (about 17 K); this is consistent with previous measurements of specific heat and susceptibility. To reveal the intrinsic quantum dynamics of the local moments of Yb, we used inelastic neutron scattering (INS) to study the spin excitations in single crystals of YbMgGaO₄ at approximately 70 mK. Constant-energy images are presented in Fig. 2a–e, which indicate the presence of diffusive magnetic excitations for all measured energies. The scattering spectral weights are spread broadly in the Brillouin zone, whereas the spectral intensities near the zone centre (that is, the Γ point) are suppressed. For a low-energy transfer of 0.3 meV, the spectral intensity

Here t is the hopping amplitude between nearest-neighbour sites.

g, Spinon Fermi surface calculated using the model described in the main text. The black arrow indicates a spinon particle–hole excitation with momentum transfer \mathbf{p} and dashed lines indicate the Brillouin zone boundaries of the conventional unit cell ($a = b = 3.40$ Å, $c = 25.12$ Å). High-symmetry points M , K and Γ are labelled by red, green and blue dots, respectively. The wave vector \mathbf{Q} is defined as $\mathbf{Q} = H\mathbf{a}^* + K\mathbf{b}^* + L\mathbf{c}^*$; a.u., arbitrary units; r.l.u., reciprocal lattice units.

is slightly more pronounced around the M points, while the broad continuum across the Brillouin zone still carries the vast majority of the spectral weight (Fig. 2a).

Figure 3a displays a contour plot of spectral intensity along the high-symmetry momentum directions (M – K – Γ – M – Γ) in energy–momentum (E – \mathbf{Q}) space. Similarly to the constant-energy images shown in Fig. 2a–e, the spectral intensity is broadly distributed in momentum for all of the energies measured. Moreover, a clear V-shaped upper bound on the excitation energy is evident near the Γ point (Fig. 3a, dotted line). The intensity of the spin excitation gradually decreases with increasing energy, and vanishes above approximately 1.5 meV. This feature is confirmed by the \mathbf{Q} scans in Fig. 4a, b and the E scans at a few given momentum points (Γ , M and K) in Fig. 4c.

The broad continuum is an immediate consequence, and strong evidence, of spinon excitations in QSLs^{1,7,12}. This differs from magnon-like excitations that would peak strongly at specific momenta in reciprocal space, with or without static magnetic order^{23,24}. In general, the spinful excitations in QSLs are carried by deconfined spinons^{1,7}. For most experimentally relevant QSLs, the spinons carry half-integer spins. One neutron-spin-flip event in an INS measurement creates an integer spin change that necessarily excites two (or more) spinons¹. Therefore, the energy transfer E and momentum transfer \mathbf{p} of the neutron are shared by two spinon excitations that are created by the neutron spin flip. According to energy–momentum conservation, we have $E(\mathbf{p}) = \omega_s(\mathbf{k}) + \omega_s(\mathbf{p} - \mathbf{k})$, where $\omega_s(\mathbf{k})$ is the spinon dispersion and \mathbf{k} is the momentum of one spinon. This relation implies the presence of an excitation continuum in the INS spectrum. The broad

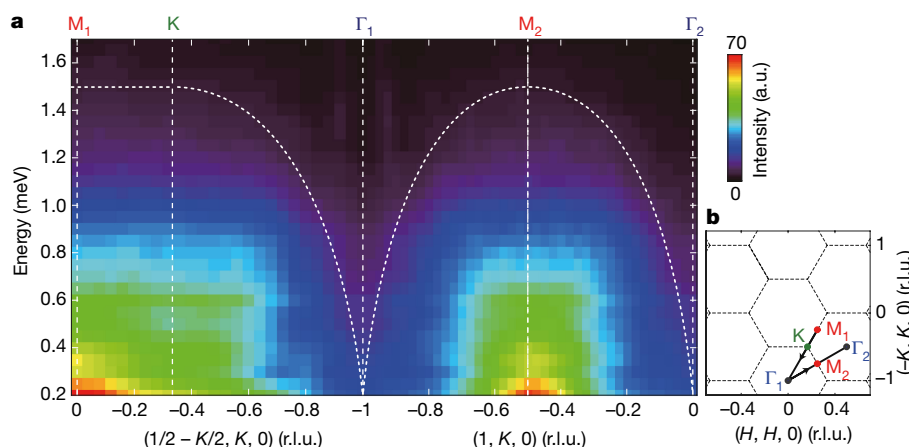


Figure 3 | Intensity of the spin-excitation spectrum along the high-symmetry momentum directions. **a**, Contour plot of the intensity along the $(1/2 - K/2, K, 0)$ and $(1, K, 0)$ directions illustrated by the black lines in **b**. The y axis represents the energy transfer. Vertical dashed lines

represent the high-symmetry points and dotted lines indicate the upper bounds on the energy of the spin excitations. **b**, Sketch of reciprocal space. Dashed lines indicate the Brillouin zone boundaries.

continua in Figs 2a–e and 3a at different energies are as expected for the continuum excitations of two spinons.

The broad neutron-scattering spectral intensity that persists to the lowest energy that we measured suggests a high density of spinon scattering states at low energies. This cannot be explained by a Dirac QSL^{2,25}, in which the spectral continuum at low energies would concentrate near a few discrete momenta that connect the Dirac cones (Methods), or by any simple gapped QSL. Because of the gap, the spectral intensity would exceed a specified energy threshold. Even if the gap was smaller than the lowest measured energy, the spinon excitations would, except under special circumstances¹⁶, occupy only one or a few discrete spots in reciprocal space at low energies²⁶, gradually expanding with increasing energy; a broad continuum at all energies and diminishing spectral weight at Γ (Fig. 3a) would not be observed. Moreover, Dirac and gapped QSLs are inconsistent with the observed low-temperature sublinear power-law behaviour of the heat capacity¹⁹. In contrast, the spinon-Fermi-surface QSL, with a high density of spinon states near the spinon Fermi surface, provides a consistent explanation for the INS results of YbMgGaO₄.

To account for these possible QSL signatures in YbMgGaO₄, we consider a minimal mean-field spinon Hamiltonian with a uniform spinon hopping on the triangular lattice. With a zero background flux for the spinons, the spinons form a large Fermi surface in the Brillouin zone (Fig. 2g). Although the anisotropic spin exchange caused by the spin-orbit coupling^{19–21} breaks the spin-rotational symmetry of this simple model, the mean-field state considered here captures the

essential properties of the spinon-Fermi-surface QSL in this system. For this spinon Fermi surface state, one neutron spin flip excites one spinon particle-hole pair across the Fermi surface. Therefore, the dynamic spin structure factor $S(\mathbf{p}, E)$, measured by INS, directly probes the spinon particle-hole excitations across the spinon Fermi surface (Fig. 2g).

For low E , a minimum momentum transfer $p_{\min} \approx E/v_F$, where v_F is the Fermi velocity, is required to excite the spinon particle-hole pairs. Therefore, the spectral intensity near the Γ point should be gradually suppressed with increasing energy, leading to an upper bound on the excitation energy near the Γ point, as is clearly observed in Fig. 3a (V-shaped dotted line). For a typical, finite E , the calculated spectrum based on the spinon particle-hole continuum is shown in Fig. 2f. This spectrum is qualitatively consistent with the experimental observation of the broad spinon continuum in reciprocal space. Finally, when E exceeds the spinon bandwidth, the single spinon particle-hole excitation process is suppressed, and the spinon excitation intensity is suppressed accordingly. This feature is consistent with the vanishing of the spectral intensity above about 1.5 meV (dotted line) seen in Figs 3a and 4c. Therefore, we propose that YbMgGaO₄ is a QSL with a spinon Fermi surface.

The spinon Fermi surface alone has a constant density of states and would give a heat capacity that depends linearly on temperature ($C_v \propto T$). To account for the $C_v \propto T^{2/3}$ behaviour in YbMgGaO₄¹⁹, we further propose that the candidate QSL is a spinon-Fermi-surface U(1) QSL, where the strong U(1) gauge fluctuation invokes a self-energy

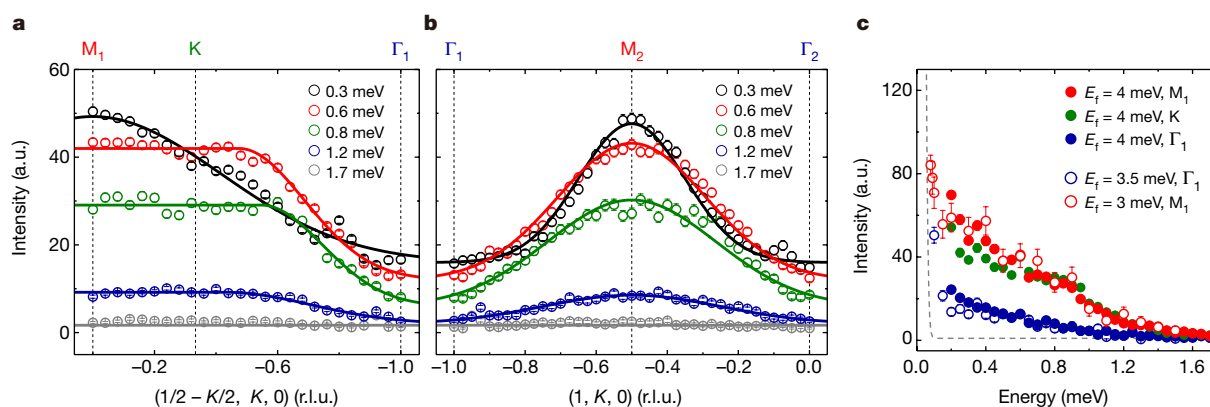


Figure 4 | Constant-energy scans along the symmetry directions and constant-Q scans at the high-symmetry points. **a**, **b**, Constant-energy scans along the $(1/2 - K/2, K, 0)$ and $(1, K, 0)$ directions. The solid lines are guides to the eye. **c**, Constant-Q scans at the M, K and Γ points with

the final energy E_f fixed at 3 meV, 3.5 meV and 4 meV, as indicated. The sharp upturn of the scattering below about 0.1 meV is due to contamination from incoherent elastic scattering at $E = 0$ meV (dashed line, $E_f = 3$ meV). Error bars, 1 s.d.

correction in the spinons, thus enhancing the low-energy density of states^{27–29}.

During the review of the Letter, a related preprint³⁰ appeared that discusses the role of the next-nearest-neighbour interaction in the formation of the QSL state in YbMgGaO₄.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 June; accepted 4 November 2016.

Published online 5 December 2016.

1. Balents, L. Spin liquids in frustrated magnets. *Nature* **464**, 199–208 (2010).
2. Wen, X.-G. Quantum orders and symmetric spin liquids. *Phys. Rev. B* **65**, 165113 (2002).
3. Kivelson, S. A., Rokhsar, D. S. & Sethna, J. P. Topology of the resonating valence-bond state: solitons and high- T_c superconductivity. *Phys. Rev. B* **35**, 8865(R)–8868(R) (1987).
4. Anderson, P. W. Resonating valence bonds: a new kind of insulator? *Mater. Res. Bull.* **8**, 153–160 (1973).
5. Anderson, P. W. The resonating valence bond state in La₂CuO₄ and superconductivity. *Science* **235**, 1196–1198 (1987).
6. Moessner, R. & Sondhi, S. L. Resonating valence bond liquid physics on the triangular lattice. *Prog. Theor. Phys.* **115** (Suppl.), 37–42 (2002).
7. Lee, P. A. An end to the drought of quantum spin liquids. *Science* **321**, 1306–1307 (2008).
8. Kurosaki, Y., Shimizu, Y., Miyagawa, K., Kanoda, K. & Saito, G. Mott transition from a spin liquid to a Fermi liquid in the spin-frustrated organic conductor κ -(ET)₂Cu₂(CN)₃. *Phys. Rev. Lett.* **95**, 177001 (2005).
9. Itou, T., Oyamada, A., Maegawa, S., Tamura, M. & Kato, R. Quantum spin liquid in the spin-1/2 triangular antiferromagnet EtMe₃Sb[Pd(dmit)₂]₂. *Phys. Rev. B* **77**, 104413 (2008).
10. Shimizu, Y., Miyagawa, K., Kanoda, K., Maesato, M. & Saito, G. Spin liquid state in an organic Mott insulator with a triangular lattice. *Phys. Rev. Lett.* **91**, 107001 (2003).
11. Coldea, R., Tennant, D. A. & Tylczynski, Z. Extended scattering continua characteristic of spin fractionalization in the two-dimensional frustrated quantum magnet Cs₂CuCl₄ observed by neutron scattering. *Phys. Rev. B* **68**, 134424 (2003).
12. Han, T.-H. *et al.* Fractionalized excitations in the spin-liquid state of a kagome-lattice antiferromagnet. *Nature* **492**, 406–410 (2012).
13. Oshikawa, M. Commensurability, excitation gap, and topology in quantum many-particle systems on a periodic lattice. *Phys. Rev. Lett.* **84**, 1535–1538 (2000).
14. Hastings, M. B. Lieb–Schultz–Mattis in higher dimensions. *Phys. Rev. B* **69**, 104431 (2004).
15. Lieb, E., Schultz, T. & Mattis, D. Two soluble models of an antiferromagnetic chain. *Ann. Phys.* **16**, 407–466 (1961).
16. Punk, M., Chowdhury, D. & Sachdev, S. Topological excitations and the dynamic structure factor of spin liquids on the kagome lattice. *Nat. Phys.* **10**, 289–293 (2014).
17. Watanabe, H., Po, H. C., Vishwanath, A. & Zaletel, M. Filling constraints for spin–orbit coupled insulators in symmorphic and nonsymmorphic crystals. *Proc. Natl Acad. Sci. USA* **112**, 14551–14556 (2015).
18. Li, Y. *et al.* Gapless quantum spin liquid ground state in the two-dimensional spin-1/2 triangular antiferromagnet YbMgGaO₄. *Sci. Rep.* **5**, 16419 (2015).
19. Li, Y. *et al.* Rare-earth triangular lattice spin liquid: a single-crystal study of YbMgGaO₄. *Phys. Rev. Lett.* **115**, 167203 (2015).
20. Li, Y.-D., Wang, X. & Chen, G. Anisotropic spin model of strong spin–orbit-coupled triangular antiferromagnets. *Phys. Rev. B* **94**, 035107 (2016).
21. Li, Y.-D., Shen, Y., Li, Y., Zhao, J. & Chen, G. The effect of spin–orbit coupling on the effective-spin correlation in YbMgGaO₄. Preprint at <http://arxiv.org/abs/1608.06445> (2016).
22. Ross, K. A., Savary, L., Gaulin, B. D. & Balents, L. Quantum excitations in quantum spin ice. *Phys. Rev. X* **1**, 021002 (2011).
23. Zhao, J. *et al.* Neutron scattering measurements of spatially anisotropic magnetic exchange interactions in semiconducting K_{0.85}Fe_{1.54}Se₂ ($T_N = 280$ K). *Phys. Rev. Lett.* **112**, 177002 (2014).
24. Wang, Q. *et al.* Magnetic ground state of FeSe. *Nat. Commun.* **7**, 12182 (2016).
25. Ran, Y., Hermele, M., Lee, P. A. & Wen, X.-G. Projected-wave-function study of the spin-1/2 Heisenberg model on the Kagomé lattice. *Phys. Rev. Lett.* **98**, 117205 (2007).
26. Sachdev, S. Kagomé- and triangular-lattice Heisenberg antiferromagnets: ordering from quantum fluctuations and quantum-disordered ground states with unconfined bosonic spinons. *Phys. Rev. B* **45**, 12377–12396 (1992).
27. Lee, S.-S. & Lee, P. A. U(1) gauge theory of the Hubbard model: spin liquid states and possible application to κ -(BEDT-TTF)₂Cu₂(CN)₃. *Phys. Rev. Lett.* **95**, 036403 (2005).
28. Motrunich, O. I. Variational study of triangular lattice spin-1/2 model with ring exchanges and spin liquid state in κ -(ET)₂Cu₂(CN)₃. *Phys. Rev. B* **72**, 045105 (2005).
29. Lee, P. A. & Nagaosa, N. Gauge theory of the normal state of high- T_c superconductors. *Phys. Rev. B* **46**, 5621–5639 (1992).
30. Paddison, J. A. M. *et al.* Continuous excitations of the triangular-lattice quantum spin liquid YbMgGaO₄. Preprint at <http://arxiv.org/abs/1607.03231> (2016).

Acknowledgements We thank D. Lee, S. Li, Y. Lu, X. Wang and, especially, J.-W. Mei for discussions, and F. Song for assistance with magnetic susceptibility measurements. This work was supported by the National Key R&D Program of the MOST of China (grant number 2016YFA0300203), the Ministry of Science and Technology of China (Program 973: 2015CB921302), and the National Natural Science Foundation of China (grant number 91421106). Y.-D.L. and G.C. were supported by the Thousand Youth Talent Program of China. Q.M.Z. was supported by the NSF of China and the Ministry of Science and Technology of China (grant number 2016YFA0300504). A portion of this research used resources at the High Flux Isotope Reactor, a DOE Office of Science User Facility operated by the Oak Ridge National Laboratory.

Author Contributions J.Z. and G.C. planned the project. Y.S., H.W. and S.S. synthesized the sample. Y.S. and Y.H. characterized the sample with the help of Y.L. and Q.Z. Y.S., H.W., S.S., B.P., Q.W., H.C.W., P.S. and J.Z. carried out the neutron experiments with experimental assistance from M.B., D.L.Q.-C., L.W.H., L.H., M.D.F. and S.M. J.Z. and Y.S. analysed the data. Y.-D.L. and G.C. provided the theoretical explanation and calculations. J.Z., G.C., Y.S. and Y.-D.L. wrote the paper. All authors provided comments on the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.Z. (zhaoj@fudan.edu.cn) or G.C. (gchen_physics@fudan.edu.cn).

Reviewer Information Nature thanks C. Rüegg and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Sample growth and characterizations. High-quality YbMgGaO₄ single crystals were synthesized using the optical floating zone technique¹⁹. A representative single crystal, which is optically transparent with mirror-like cleaved surfaces, is shown in Extended Data Fig. 1a. Our X-ray diffraction (XRD) measurements revealed that all of the reflections from the cleaved surface could be indexed by (0, 0, *L*) peaks of triangular YbMgGaO₄; no impurity phases were observed (Extended Data Fig. 1b). The full-width at half-maximum (FWHM) of the rocking curve of the (0, 0, 18) peak was about 0.009°, indicating an extremely high crystallization quality (Extended Data Fig. 1c). This was confirmed by the sharp and clear diffraction spots in the X-ray Laue pattern (Extended Data Fig. 1d). Powder XRD patterns on ground single crystals also revealed no indication of impurity phases (Extended Data Fig. 1e). The Rietveld refinements³¹ confirm that the XRD pattern can be described by the $R\bar{3}m$ space group. The refined structural parameters are given in Extended Data Table 1. These results suggested that the YbMgGaO₄ single crystal possessed a perfect triangular lattice with no detectable impurities. This is consistent with previous measurements that have demonstrated that the impurity/isolated spins are less than 0.04% in similar samples^{18,19}. Although the Mg/Ga site disorder in the non-magnetic layers does not directly affect the exchange interaction between the Yb local moments, it may have an indirect effect and could lead to some exchange disorder. It seems that this disorder is not significant, because no signs of spin freezing were observed. A QSL is often stable against weak local perturbations, provided that the perturbation is irrelevant or not significant. Therefore, if a QSL is realized as the ground state for YbMgGaO₄, then the possible exchange disorder will not destabilize this state if the disorder strength is not significant.

In addition, the field dependence of magnetization in our single crystal displayed a linear behaviour above 12 T (Extended Data Fig. 1f), indicative of a fully polarized state. The Van Vleck susceptibility extracted from the linear-field-dependent magnetization data was subtracted in the inset of Fig. 1c. **Neutron scattering experiments.** INS measurements were carried out on the ThALES cold triple-axis spectrometer at the Institut Laue-Langevin, Grenoble, France, and at the FLEXX cold triple-axis spectrometer in the BER-II reactor at Helmholtz-Zentrum Berlin, Germany³². For the ThALES experiment, silicon (111) was used as a monochromator and analyser; the final neutron energies were fixed at $E_f = 3$ meV (energy resolution of about 0.05 meV), $E_f = 3.5$ meV (energy resolution of about 0.08 meV) or $E_f = 4$ meV (energy resolution of about 0.1 meV). For the FLEXX experiment, pyrolytic graphite (002) was used as a monochromator and analyser. Contamination from higher-order neutrons was eliminated through a velocity selector installed in the front of the monochromator. The final neutron energy was fixed at $E_f = 3.5$ meV (energy resolution of about 0.09 meV). Three (six) pieces of single crystals with total a mass of about 5 g (19 g) were coaligned in the (HK0) scattering plane for the ThALES (FLEXX) experiment. The FWHM of the rocking curve of the coaligned crystals for the ThALES and FLEXX experiments were approximately 0.95° and 0.92°, respectively. The elastic neutron scattering experiment was carried out at the WAND neutron diffractometer at the High Flux Isotope Reactor, Oak Ridge National Laboratory, USA; one single crystal was used for the experiment, with the incident wavelength $\lambda = 1.488$ Å (Extended Data Fig. 2). For the low-temperature experiments, a dilution insert for the standard ⁴He cryostat was used to reach temperatures down to around 30–70 mK.

Because of the non-uniform shape of the single crystal, the relatively large sample volume and the extremely broad spin-excitation spectrum, the neutron beam self-attenuation (by the sample) may require consideration. In most cases the self-attenuation is dependent on only the distance traversed by the neutrons through the sample. We observed the self-attenuation effect in an elastic incoherent scattering image of our sample at 20 K, which exhibited an anisotropic intensity distribution (Extended Data Fig. 3a). The self-attenuation effect was also observed in the raw constant-energy images (Extended Data Fig. 3b–f), which were shown to be anisotropic, with slightly higher intensities occurring at approximately the same direction as that observed in the elastic incoherent scattering images. The self-attenuation can be corrected by normalizing the data with the elastic incoherent scattering image; that is, the elastic incoherent scattering intensity, which is dependent on the sample position (ω) and scattering angle (2θ), is converted to a linear attenuation correction factor for the scattering images measured at different energies. The normalized constant-energy images are presented in Fig. 2a–e, revealing a nearly isotropic intensity distribution.

Extended Data Fig. 4 shows the spin excitation spectrum at 20 K, which is broadened and weakened compared with that at 70 mK (discussed below).

Spinon Fermi surface and dynamic spin structure factor. Here we explain the spinon mean-field state that is used to explain the dynamic spin structure factor of the neutron scattering experiments. As we proposed in the main text, a QSL with a spinon Fermi surface gives a compatible explanation for the INS results for YbMgGaO₄.

To describe the candidate spinon-Fermi-surface QSL state in YbMgGaO₄, we formally express the Yb³⁺ effective spin as the bilinear combination of the fermionic spinon with spin $\mathbf{S}_i = \sum_{\alpha\beta} \frac{1}{2} f_{i\alpha}^\dagger \boldsymbol{\sigma}_{\alpha\beta} f_{i\beta}$ and a Hilbert space constraint $\sum_{\alpha} f_{i\alpha}^\dagger f_{i\alpha} = 1$, where $\boldsymbol{\sigma}_{\alpha\beta}$ is a vector whose three components are the Pauli matrices and $f_{i\alpha}^\dagger (f_{i\alpha})$ creates (annihilates) a spinon with spin $\alpha = \uparrow, \downarrow$ at site *i*. For the QSL with a spinon Fermi surface, we propose a minimal mean-field Hamiltonian H_{MFT} for the spinons on the triangular lattice. We consider a uniform spinon hopping with a zero background flux:

$$H_{\text{MFT}} = -t \sum_{\langle ij \rangle} (f_{i\alpha}^\dagger f_{j\alpha} + \text{h.c.}) - \mu \sum_i f_{i\alpha}^\dagger f_{i\alpha} \quad (1)$$

where *t* is the mean-field parameter, which represents the hopping amplitude between nearest-neighbour sites. The chemical potential μ is included to impose the Hilbert space constraint on average. Here, we have treated the spinons freely by neglecting the gauge fluctuations. This mean-field state gives a single spinon dispersion

$$\omega_{\mathbf{k}} = -t \sum_{\{\mathbf{a}_i\}} \cos(\mathbf{k} \cdot \mathbf{a}_i) - \mu$$

where $\{\mathbf{a}_i\}$ are six nearest-neighbour vectors of the triangular lattice. Owing to the Hilbert space constraint, the spinon band is half-filled, leading to a large Fermi surface in the Brillouin zone (Extended Data Fig. 5a).

INS measures the dynamic spin structure factor

$$\begin{aligned} S(\mathbf{p}, E) &= \frac{1}{N} \sum_{i,j} e^{i\mathbf{p} \cdot (\mathbf{r}_i - \mathbf{r}_j)} \int e^{-iEt} \langle \mathbf{S}_i^\dagger(t) \cdot \mathbf{S}_j^\dagger(0) \rangle dt \\ &= \sum_n \delta(E - [E_n(\mathbf{p}) - E_0]) \left| \langle n | \mathbf{S}_p^\dagger | \Omega \rangle \right|^2 \end{aligned} \quad (2)$$

where *N* is total number of lattice sites, the summation goes over all eigenstates, $|\Omega\rangle$ refers to the spinon ground state with the spinons filling the Fermi sea, E_0 is the energy of the ground state and $E_n(\mathbf{p})$ is the energy of the *n*th excited state with momentum \mathbf{p} . In the actual calculation, owing to the energy resolution of the experiments, the δ function is taken to have a broadening:

$$\delta(E - \epsilon) = \frac{\eta/\pi}{(E - \epsilon)^2 + \eta^2}$$

where η is the broadening and ϵ is the measured energy. Because $\mathbf{S}_p^\dagger = \sum_{\mathbf{k}} f_{\mathbf{k}+\mathbf{p}}^\dagger f_{\mathbf{k}}$, the summation in equation (2) would be over all possible spin-1 excited states that are characterized by one spinon particle-hole pair crossing the spinon Fermi surface (Fig. 2g) with a total momentum \mathbf{p} and a total energy *E*. As we show in Fig. 2f and Extended Data Fig. 5b, and discuss in the main text, this spinon-Fermi-surface QSL state gives the three crucial features of the INS results: (1) the broad continuum that covers the large portion of the Brillouin zone; (2) the broad continuum persisting from the lowest energy transfer to the highest energy transfer; and (3) the clear upper excitation edge near the Γ point.

In our calculation of Fig. 2f and Extended Data Fig. 5b, we choose the lattice size to be 40×40 and $\eta = 1.2t$, in accordance with the energy and momentum resolution of the instruments. The energy scale of Fig. 2f is set to be 7.5*t*.

Here we explain the details of the dynamic spin structure factor in Fig. 2f and Extended Data Fig. 5b, based on the particle-hole excitation of the spinon Fermi surface. For an infinitesimal energy transfer, the neutrons simply probe the spinon Fermi surface. Because the spinon particle and hole can be excited anywhere near the Fermi surface, the neutron spectral intensity appears from $\mathbf{p} = \mathbf{0}$ to $\mathbf{p} = 2\mathbf{k}_F$, where \mathbf{k}_F is the Fermi wavevector. Because $|2\mathbf{k}_F|$ already exceeds the first Brillouin zone, the neutron spectral intensity then covers the whole Brillouin zone including the Γ point. For a small but finite *E*, as we explain in the main text, a minimal momentum transfer $p_{\text{min}} \approx E/v_F$ is required to excite the spinon particle-hole pairs. Therefore, the spectral intensity gradually moves away from the Γ point as *E* increases. Because it is always possible to excite the spinon particle-hole pair with the momenta near the zone boundary, the spectral intensity is not greatly affected at the zone boundary as *E* increases. Thus, the broad continuum continues to cover a large portion of the Brillouin zone at a finite *E*.

With the free spinon mean-field model H_{MFT} , we further calculate the spectral weight along the energy direction for fixed momenta. The discrepancy between the theoretical results in Extended Data Fig. 5d and the experimental results in Extended Data Fig. 5e occurs at low energies. We attribute this low-energy discrepancy to the fact that the free spinon theory ignores the gauge fluctuations. The enhancement of the low-energy spectral weight compared to the free spinon results is then identified as possible evidence of strong gauge fluctuations in the system; we elaborate on this in the following discussion of the heat capacity behaviour.

To account for the heat capacity behaviour, we suggest that the candidate QSL is a spinon-Fermi-surface U(1) QSL. Here we elaborate on this point and discuss the U(1) gauge fluctuation of this state in detail. The stability of U(1) QSLs with a spinon Fermi surface against the spinon confinement has been addressed extensively^{33,34}. It was proposed and understood that the large densities of gapless fermionic spinons on the spinon Fermi surface help to suppress the instanton events of the compact U(1) gauge field in two-dimensional U(1) QSLs^{33,34}. The proliferation of instanton events is the cause of the gauge confinement of a U(1) lattice gauge theory for a U(1) QSL without gapless spinons³⁵. Because the instanton event is suppressed here, the compactness of the U(1) gauge field is no longer an issue, and the low-energy property of our U(1) QSL is then described by gapless fermionic spinons coupled with a non-compact U(1) gauge field^{27,33,36}. Owing to the coupling to the gapless spinons, the U(1) gauge photon is overly Landau damped and becomes very soft. The soft gauge photon strongly scatters the fermionic spinons, provides a self-energy correction to the Green's function of the spinon, and thus makes the quasi-particle weight for spinons equal to zero^{27,29,33,36}. The resulting spinon non-Fermi liquid state has an enhanced density of low-energy spinon states that results in a sublinear power-law temperature dependence for the low-temperature heat capacity^{27,29,33,36}. In addition to the heat capacity behaviour, we find that, owing to the spinon-gauge coupling and the U(1) gauge fluctuation, the enhanced density of the low-energy spinon states is consistent with the enhanced spectral intensities at low energies for the fixed momenta in Extended Data Fig. 5e.

The stability of the spinon non-Fermi liquid against spinon pairing has also been considered theoretically³⁷. When the spinons pair up, similar to the Cooper pairing of electrons in a superconductor, the continuous part of the U(1) gauge field becomes massive, owing to the Anderson-Higgs mechanism, leaving the Z_2 part of the gauge field unaffected. The resulting state from the spinon pairing of a spinon-Fermi-surface U(1) QSL is a Z_2 QSL. Such a spinon pairing scenario was proposed to account for the very low-temperature behaviour of the organic spin liquid κ -(BEDT-TTF)₂Cu₂(CN)₃ (ref. 38). However, for YbMgGaO₄, we do not find any evidence of spinon pairings in either thermodynamic or spectroscopic measurements. Although the INS measurement might be constrained by the energy resolution, the thermodynamic measurement did not find any suppression of the density of states down to the lowest temperatures^{18,19}. If the spinon pairing instability occurs for YbMgGaO₄, it must be at a much lower temperature or energy scale than those of current and previous experiments^{18,19}. In any case, the presence of a spinon Fermi surface is the precondition for any spinon pairing instability.

We now discuss the finite-temperature thermal effect of the QSL. For the spinon-Fermi-surface U(1) QSL in two dimensions, there is no line-like object in the excitation spectrum. Therefore, as the temperature is increased from this QSL ground state, there is no thermal phase transition caused by proliferating any extended line-like excitations. Moreover, the spinon-Fermi-surface U(1) QSL is not characterized by any symmetry. Consequently, there is no symmetry-breaking transition as the temperature is increased. The absence of the thermal phase transition is consistent with what has been observed in YbMgGaO₄. As the temperature is increased from the $T = 0$ K ground state of the QSL, the system involves more thermal superposition of excited states and gradually loses its quantum coherence. A temperature of 20 K is approximately the energy scale of the spin excitation bandwidth, which sets the interaction energy scale between the Yb local moments. At this temperature, the correlation between the local moments cannot be ignored. Its consequence is the diffusive feature in the INS spectrum. This is consistent with our data measured

at 20 K shown in Extended Data Fig. 4a, b, in which the spectral weight becomes more diffusive.

Finally, we comment on the weak spectral peak at the M points at low energies (Figs 2a, 3a). This non-generic feature of the neutron spectrum is not obtained in the theoretical calculation using the minimal spinon mean-field model in equation (1). This is because we did not include the effect of anisotropic spin interaction, which would break the spin rotational symmetry of equation (1). In the strong anisotropic limit, the generic spin interaction for the Yb local moments favours stripe-like magnetic order, with the wavevectors at the M points²⁰. In a recent calculation, it was shown that the anisotropic spin interaction enhances the spin correlation at the M points²¹. Despite the presence of a weak peak at M, the vast majority of the spectral weight is still dominated by a broad continuum across the Brillouin zone at the lowest energies measured (Figs 2a, 3a).

Dynamic spin structure factor of a Dirac QSL. As a comparison with the spinon-Fermi-surface QSL, we carry out the same calculation for the spinon mean-field Hamiltonian with a background π flux through each unit cell. This choice of the background flux gives a Dirac U(1) QSL. We fix the gauge according to the hopping parameters that are specified in Extended Data Fig. 6a. The spinon band structure of this mean-field Hamiltonian is

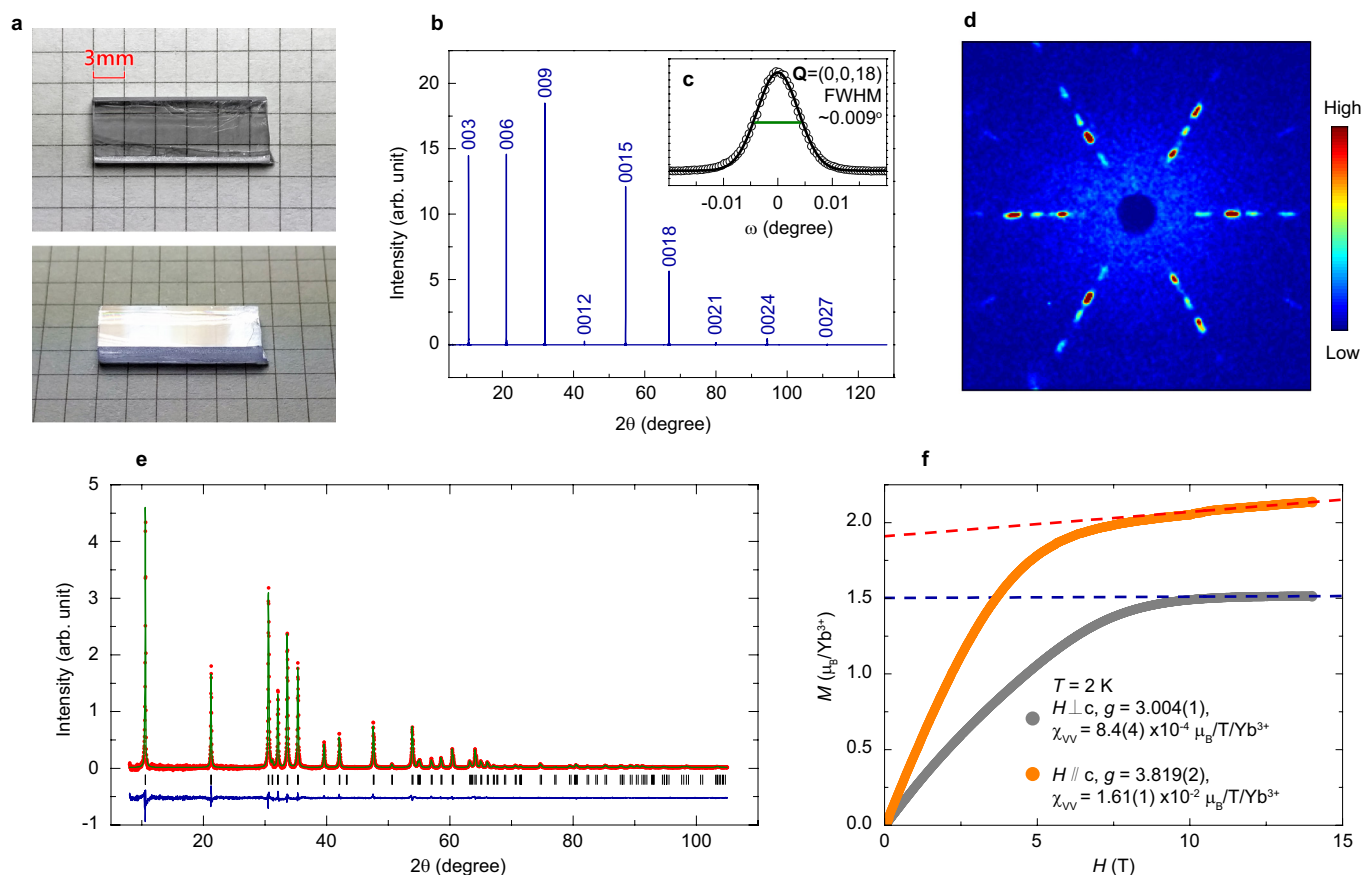
$$\omega_{\mathbf{k}} = \pm \sqrt{2} t \sqrt{3 + \cos(2k_x) + 2\sin(k_x)\sin(\sqrt{3}k_y)}$$

where we have set the lattice constant to unity. We observe two Dirac nodes at $\mathbf{k} = (\pm\pi/2, \mp\pi/(2\sqrt{3}))$ (Extended Data Fig. 6b); the spinon Fermi energy is right at the Dirac nodes.

At low energies, the only spin-1 excited states involve either an intra-Dirac-cone spinon particle-hole pair or an inter-Dirac-cone particle-hole pair. Therefore, the spectral intensity of the dynamic spin structure factor should be concentrated at the momentum transfer that corresponds to the intra-Dirac-cone and the inter-Dirac-cone processes. As shown Extended Data Fig. 6c, d, the dynamic spin structure factor at low energies is peaked at the Γ point, the $M = (0, 2\pi/\sqrt{3})$ point and the symmetry-equivalent momentum points. This result differs from the broad continuum that is observed in the experiment. Therefore, the π -flux state is inconsistent with the experimental data, as are other Dirac spin liquids.

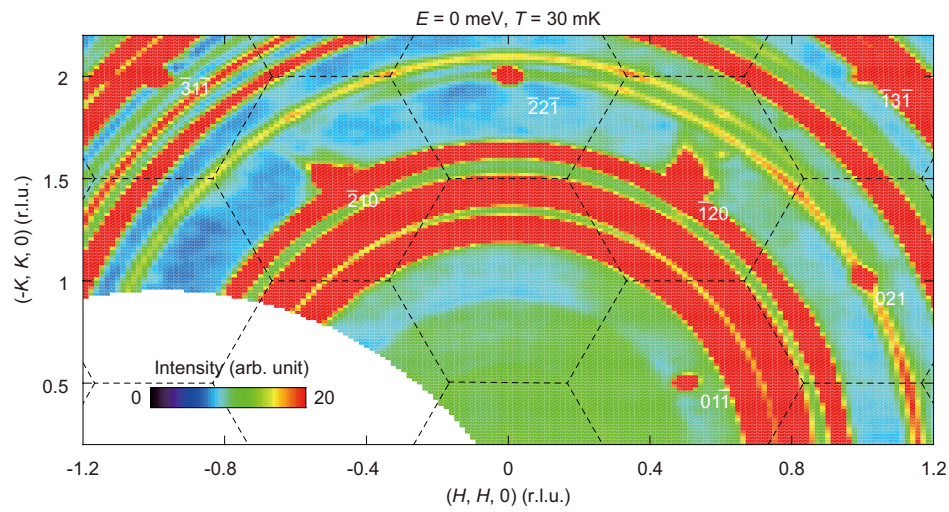
Data availability. The data that support the findings of this study are available from the corresponding author on reasonable request.

- Rodríguez-Carvajal, J. Recent advances in magnetic structure determination by neutron powder diffraction. *Physica B* **192**, 55–69 (1993).
- Le, M. D. *et al.* Gains from the upgrade of the cold neutron triple-axis spectrometer FLEXX at the BER-II reactor. *Nucl. Instrum. Methods A* **729**, 220–226 (2013).
- Lee, S. S. Stability of the U(1) spin liquid with a spinon Fermi surface in 2 + 1 dimensions. *Phys. Rev. B* **78**, 085129 (2008).
- Hermele, M. *et al.* Stability of U(1) spin liquids in two dimensions. *Phys. Rev. B* **70**, 214437 (2004).
- Polyakov, A. M. *Gauge Fields and Strings* Ch. 4 (Harwood Academic, 1987).
- Lee, S. S. Low-energy effective theory of Fermi surface coupled with U(1) gauge field in 2 + 1 dimensions. *Phys. Rev. B* **80**, 165102 (2009).
- Metlitski, M. A., Mross, D. F., Sachdev, S. & Senthil, T. Cooper pairing in non-Fermi liquids. *Phys. Rev. B* **91**, 115111 (2015).
- Lee, S. S., Lee, P. A. & Senthil, T. Amperean pairing instability in the U(1) spin liquid state with Fermi surface and application to κ -(BEDT-TTF)₂Cu₂(CN)₃. *Phys. Rev. Lett.* **98**, 067006 (2007).



Extended Data Figure 1 | Photographs, XRD patterns and field dependence of the magnetization of YbMgGaO₄. **a**, Photographs of a representative YbMgGaO₄ single crystal. **b**, XRD pattern of a YbMgGaO₄ single crystal from the cleaved surface. **c**, Rocking curve of the (0, 0, 18) peak. The horizontal bar indicates the instrumental resolution. **d**, Laue pattern of the YbMgGaO₄ single crystal viewed from

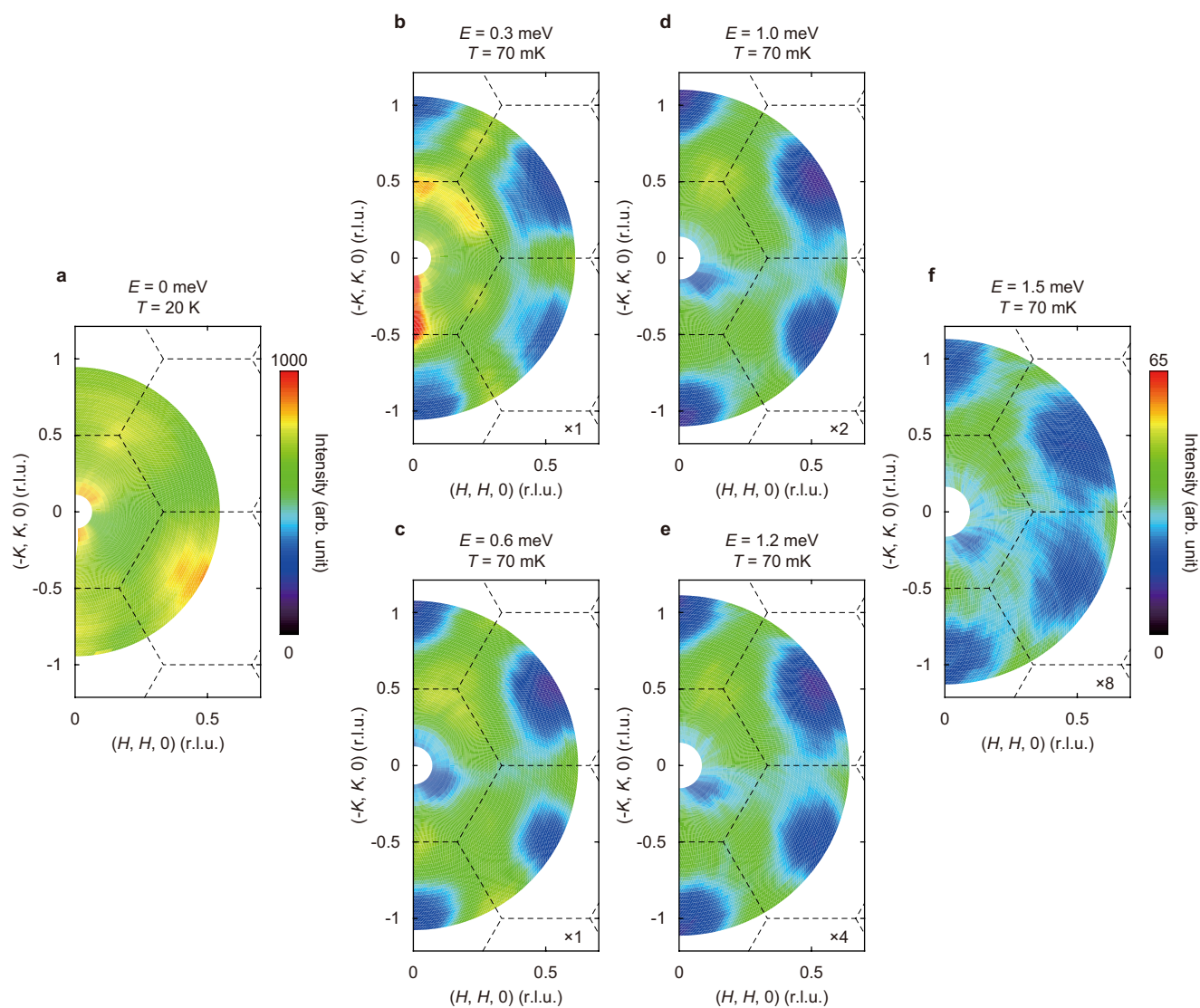
the *c* axis. **e**, Observed (red) and calculated (green) XRD diffraction intensities of ground single crystals. The X-ray has a wavelength of 1.54 Å. The blue curve indicates the difference between the observed and calculated intensities. **f**, Magnetic field dependence of magnetization at *T* = 2 K. Fitted *g* factors and Van Vleck susceptibilities χ_{VV} are shown (μ_B is the Bohr magneton). The dashed lines are linear fits above 12 T.



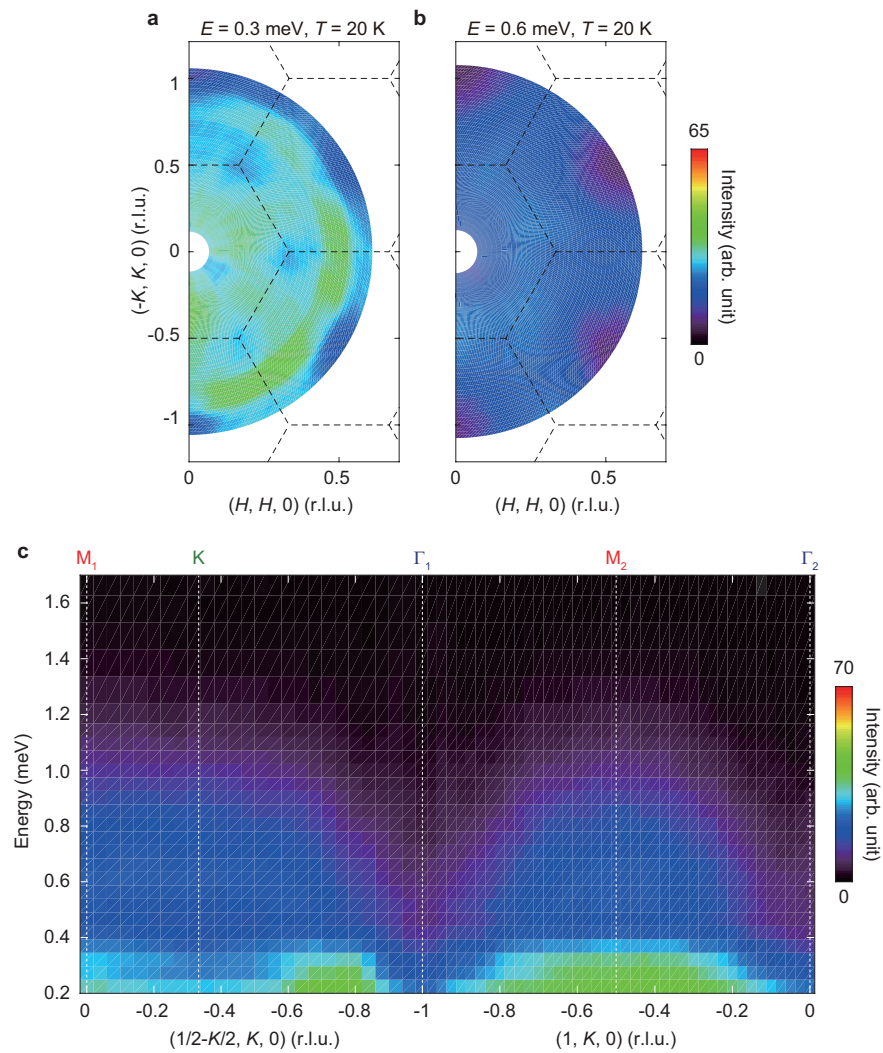
Extended Data Figure 2 | Elastic neutron scattering measurements.

Elastic neutron scattering map in the $(HK0)$ plane at 30 mK. No magnetic Bragg peaks are observed. The ring-like pattern is due to scattering from the polycrystalline Cu and Al sample holder. Because of the very large

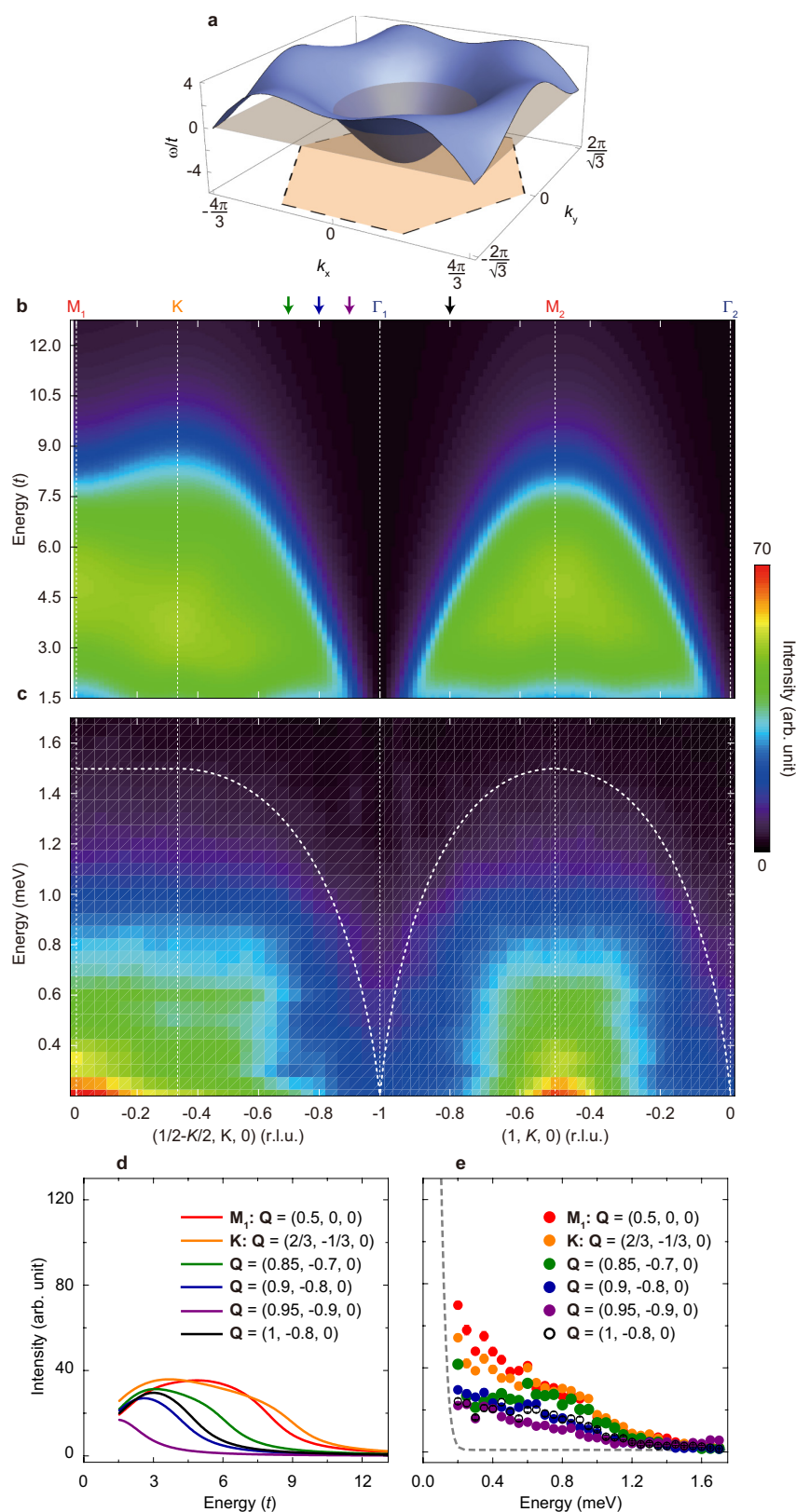
c -axis lattice constant and a small tilt of the scattering plane, some of the tails of the nuclear Bragg peaks for $L = \pm 1$ can be also seen. Dashed lines indicate the Brillouin zone boundaries.



Extended Data Figure 3 | Correction of neutron beam self-attenuation. **a**, Elastic incoherent scattering image at 20 K . **b–f**, Raw constant-energy images at 70 mK and at the indicated energies. The scattering intensities in **d**, **e** and **f** have been multiplied by 2, 4 and 8, respectively, for clarity. Dashed lines indicate the Brillouin zone boundaries.



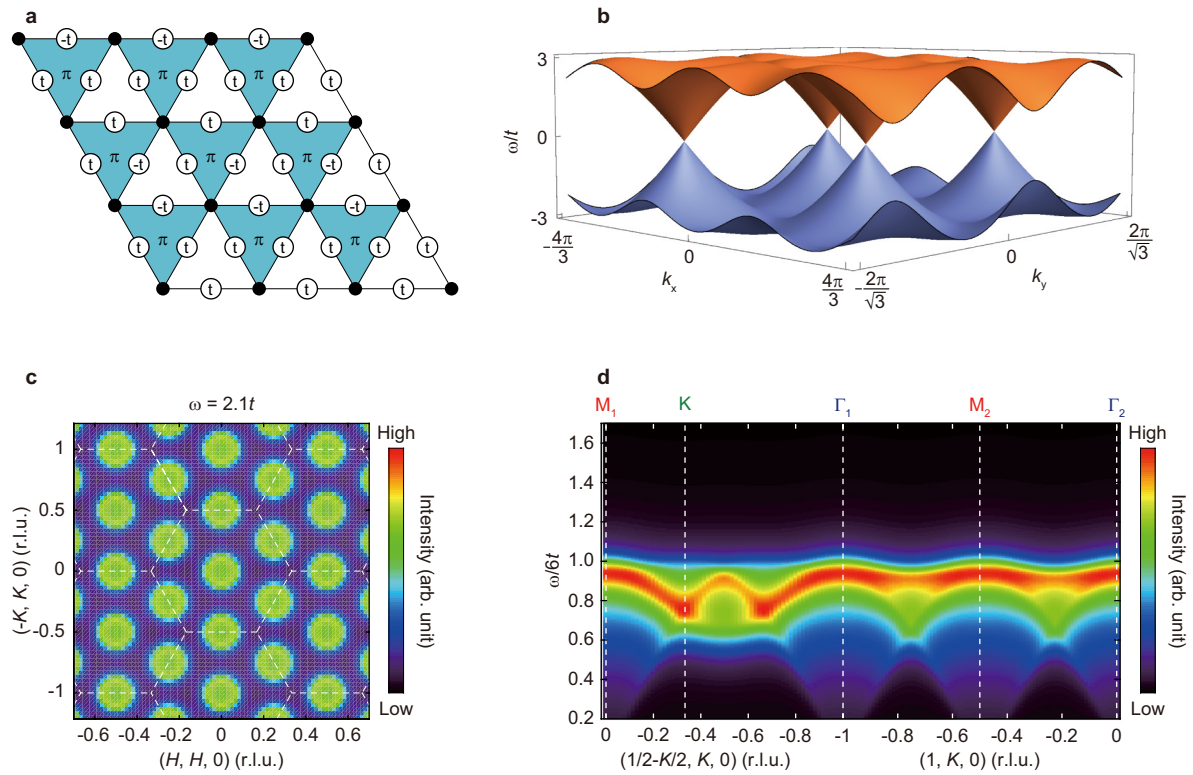
Extended Data Figure 4 | Additional neutron scattering data at 20 K. **a, b,** Constant-energy images at 0.3 meV (**a**) and 0.6 meV (**b**) at 20 K. **c,** Intensity contour plot of the spin excitation spectrum along the high-symmetry momentum directions at 20 K. The scattering is broadened and weakened compared with that at 70 mK.



Extended Data Figure 5 | Calculation of the zero-flux Hamiltonian.

a, Spinon dispersion ω_k of the zero-flux Hamiltonian. The grey plane marks the Fermi level at $\omega = 0$; its intersection with the band gives the Fermi surface. The light orange hexagon represents the projection of the first Brillouin zone. The maximum of ω_k is $3t$ and the minimum is $-6t$, providing a bandwidth of $9t$. **b**, Calculated dynamic spin structure factor along high-symmetry directions. A reciprocal lattice unit (r.l.u.) is used

here, which is obtained using $H = k_x/(4\pi) - \sqrt{3}k_y/(4\pi)$ and $K = k_x/(4\pi) + \sqrt{3}k_y/(4\pi)$. **c**, Measured spin excitation spectrum along high-symmetry directions at 70 mK. **d**, Calculated energy dispersion at the indicated momenta (marked by arrows in **b**). **e**, Measured constant- Q scans at the indicated momenta. The dashed line is the incoherent elastic line for $E_f = 4$ meV.



Extended Data Figure 6 | Calculation of the π -flux Hamiltonian. **a**, Flux pattern and real nearest-neighbour hoppings on the triangular lattice. In the figure, ‘ $+t$ ’ denotes $t_{ij} = t_{ji} = t$ and ‘ $-t$ ’ denotes $t_{ij} = t_{ji} = -t$; ‘ π ’ denotes triangles that are threaded by a π flux. **b**, Spinon band structure of the π -flux Hamiltonian. The two bands are particle-hole related, both with bandwidths of $3t$. **c**, Calculated momentum dependence of the dynamic

spin structure factor at low energy $\omega = 2.1t$. Strong peaks can be distinguished at the Γ point, the $M = (0, 2\pi/\sqrt{3})$ point ($(1/2, -1/2)$ in r.l.u.) and equivalent positions. White dashed lines denote the zone boundaries. **d**, Calculated dynamic spin structure factor along high-symmetry points with $\eta = 0.3t$.

Extended Data Table 1 | Refined structural parameters for YbMgGaO₄ at room temperature

	a (Å)	3.40125 (1)
	c (Å)	25.10632 (16)
Yb	B_{11} (Å ²)	0.1332 (18)
	B_{33} (Å ²)	0.00204 (3)
Mg	z	0.21378 (6)
	B_{11} (Å ²)	0.131 (4)
	B_{33} (Å ²)	0.00161 (6)
Ga	z	0.21378 (6)
	B_{11} (Å ²)	1.131 (4)
	B_{33} (Å ²)	0.00161 (6)
O1	z	0.28887 (19)
	B_{11} (Å ²)	0.107 (9)
	B_{33} (Å ²)	0.00226 (17)
O2	z	0.12884 (17)
	B_{11} (Å ²)	0.137 (9)
	B_{33} (Å ²)	0.00089 (18)
	R_p	1.18
	wRp	1.81
	χ^2	2.25

Space group: $R\bar{3}m$ (number 166). Atomic positions: Yb, 3a (0, 0, 0); Mg, 6c (0, 0, z); Ga, 6c (0, 0, z); O, 6c (0, 0, z). B_{ij} , Debye–Waller factor; R_p , profile factor; wRp , weighted profile factor.

Self-assembly of tetravalent Goldberg polyhedra from 144 small components

Daishi Fujita^{1,2,3}, Yoshihiro Ueda^{1,3}, Sota Sato^{4,5}, Nobuhiro Mizuno⁶, Takashi Kumasaka⁶ & Makoto Fujita^{1,3}

Rational control of the self-assembly of large structures is one of the key challenges in chemistry^{1–9}, and is believed to become increasingly difficult and ultimately impossible as the number of components involved increases. So far, it has not been possible to design a self-assembled discrete molecule made up of more than 100 components. Such molecules—for example, spherical virus capsids¹⁰—are prevalent in nature, which suggests that the difficulty in designing these very large self-assembled molecules is due to a lack of understanding of the underlying design principles. For example, the targeted assembly of a series of large spherical structures containing up to 30 palladium ions coordinated by up to 60 bent organic ligands^{11–16} was achieved by considering their topologies¹⁷. Here we report the self-assembly of a spherical structure that also contains 30 palladium ions and 60 bent ligands, but belongs to a shape family that has not previously been observed experimentally¹⁷. The new structure consists of a combination of 8 triangles and 24 squares, and has the symmetry of a tetravalent Goldberg polyhedron^{18,19}. Platonic and Archimedean solids have previously been prepared through self-assembly, as have trivalent Goldberg polyhedra, which occur naturally in the form of virus capsids²⁰ and fullerenes²¹. But tetravalent Goldberg polyhedra have not previously been reported at the molecular level, although their topologies have been predicted using graph theory. We use graph theory to predict the self-assembly of even larger tetravalent Goldberg polyhedra, which should be more stable, enabling another member of this polyhedron family to be assembled from 144 components: 48 palladium ions and 96 bent ligands.

It has been shown¹² that the structures of giant M_nL_{2n} assemblies (consisting of n ions and $2n$ ligands) are geometrically restricted, in that the assemblies need to be roughly spherical, isotropic structures to minimize their surface energy. If we assume the formation of roughly spherical, regular or semi-regular polyhedra, then the possible values of n for M_nL_{2n} assemblies are limited to 6, 12, 24, 30 and 60 (refs 4, 11). On the basis of this geometric restriction, we have been able to target these structures specifically, avoiding the many possible assemblies that might form for other n . We have previously reported the self-assembly of the M_6L_{12} octahedron¹³, the $M_{12}L_{24}$ cuboctahedron¹⁴, the $M_{24}L_{48}$ rhombicuboctahedron¹⁵ and the $M_{30}L_{60}$ icosidodecahedron¹⁶ (Fig. 1a). While dedicating effort towards the self-assembly of the next target—the $M_{60}L_{120}$ rhombicosidodecahedron—we unexpectedly encountered an ‘undefined’ polyhedron with $M_{30}L_{60}$ composition and the topology depicted in Fig. 2. This metal complex is a seemingly isotropic polyhedron consisting of 8 triangles and 24 squares, clearly different from the isostructural $M_{30}L_{60}$ icosidodecahedron and with a high-symmetry topology that does not belong to that of the Archimedean solids. This polyhedron does not have a mirror plane and, unlike the Archimedean solids, features molecular chirality defined by its topology (Fig. 2d).

We observed this $M_{30}L_{60}$ framework when the experimental procedures were carried out under conditions typical for targeting self-assembly

of standard M_nL_{2n} complexes^{4,11}, except that selenophene-cored bipyridyl-type ligand **1** was used instead of an organic bipyridyl ligand (Fig. 1b). The bend angle (θ) of ligand **1** is 152° , only 3° larger than that of thiophene-cored ligand **2** ($\theta = 149^\circ$), which selectively assembles into the $M_{24}L_{48}$ rhombicuboctahedron upon palladium(II) coordination¹⁵. A small difference in the bend angles of two ligands can critically switch the resultant self-assembled structure, as has been observed for the $M_{12}L_{24}$ -to- $M_{24}L_{48}$ transition at around $\theta = 131^\circ$ – 134° (ref. 22).

In addition to the observation of a simple ^1H nuclear magnetic resonance (^1H NMR) spectrum (Methods and Extended Data Fig. 1b), X-ray crystallographic analysis showed the three-dimensional coordination geometry of an $M_{30}L_{60}$ complex (**3**) with initially undefined topology (Fig. 2). As for the other large M_nL_{2n} polyhedra, the diffraction data collected at the synchrotron X-ray facility resembled data obtained for protein crystals rather than for conventional metal complexes. Therefore, the model building was carried out using ligand/solvent model alignment on the $2F_o - F_c$ map and on the electron density map obtained using the maximum entropy method (MEM)^{23,24}, in a manner analogous to that used for protein crystallography (Fig. 2b). For the M_nL_{2n} polyhedra, the data resolution is sufficiently high to be able to study the overall topology of the structure on the basis of the cloudless electron density map¹⁶. The refined three-dimensional structure of **3** (Fig. 2c) consists of 30 palladium(II) ions and 60 organic

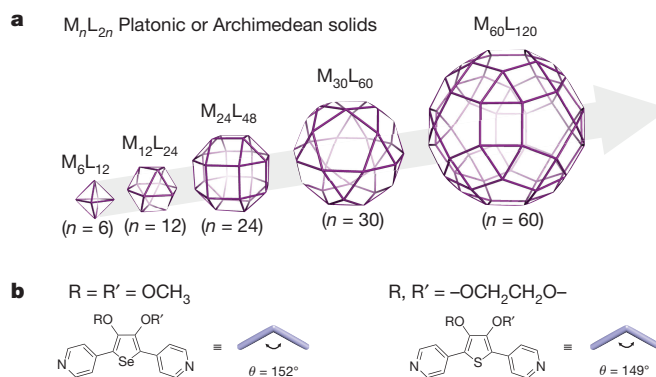


Figure 1 | M_nL_{2n} -type polyhedral metal-organic ligand complexes.

a, Schematic representation of M_nL_{2n} complexes with the symmetry of Platonic or Archimedean solids. Each vertex represents a metal ion centre and each edge represents an organic ligand. Under the prerequisite that each vertex connects to four edges (because the metal ion used in this system (palladium(II)) has a square planar coordination geometry), only five structures are allowed. M_nL_{2n} complexes with $n = 6, 12, 24$ and 30 have previously been synthesized. **b**, Molecular structures of the organic ligands. The bend angle of the ligand θ dictates the final self-assembled product. A very small change in θ can result in very different final products.

¹Department of Applied Chemistry, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. ²PRESTO (Precursory Research for Embryonic Science and Technology), Japan Science and Technology Agency, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. ³ACCEL, Japan Science and Technology Agency, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. ⁴Advanced Institute for Materials Research, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan. ⁵ERATO, Japan Science and Technology Agency, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan. ⁶Japan Synchrotron Radiation Research Institute (JASRI), Spring-8, 1-1-1 Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5198, Japan.

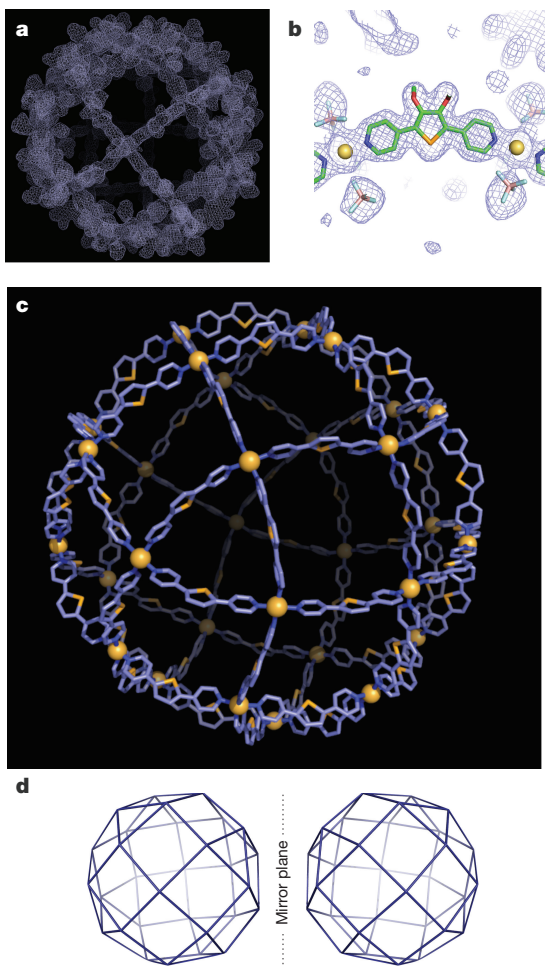


Figure 2 | X-ray crystallographic analysis of self-assembled product 3. **a**, MEM (maximum entropy method) electron density map (0.6 electrons per \AA^3) of $M_{30}L_{60}$ complex **3**. Supplementary Video 1 shows a rotating electron density map. **b**, Enlarged view around a ligand in $M_{30}L_{60}$ with its MEM electron density map. The modelled ligand framework, metal ions and counter ions (BF_4^-) agree well with the observed electron density. **c**, Entire view of the X-ray crystallographic structure of $M_{30}L_{60}$. Methoxy substituents and counter ions are omitted for clarity. Space group, $Pbca$; lattice parameters, $a = 73.0 \text{ \AA}$, $b = 73.2 \text{ \AA}$ and $c = 143.4 \text{ \AA}$; data resolution, 1.95 \AA ; $R(I > 2\sigma(I))$, 0.2049; $R_{\text{free}}(I > 2\sigma(I))$, 0.2391. **d**, Simplified image of the obtained structure. Each vertex represents a metal ion centre and each edge represents an organic ligand. The structure has polyhedral chirality; a pair of left- and right-hand forms is shown.

bidentate ligands, and forms a roughly spherical polyhedron that is described by a combination of 8 triangles and 24 squares.

The symmetry and topology of polyhedron **3** can be understood by extending the description of Goldberg polyhedra¹⁸. Goldberg polyhedra, as first described in 1937, are convex polyhedra made up of hexagons and pentagons (Fig. 3a), and provide a theoretical foundation with which to describe the topology of fullerenes²¹ and some virus capsids²⁰. In this family of Goldberg polyhedra, 12 pentagons are evenly distributed within a hexagonal honeycomb sheet that fully circumscribes a sphere. Each polyhedron is defined by the relative position of the closest pentagons using two indices, h and k , and the notation $G(h, k)$. Triangulation (T) numbers are defined by the equation $T = h^2 + hk + k^2$, where h and k are non-negative integers and so $T = 1, 3, 4, 7, 9, 12, 13, 16, \dots$ (Fig. 3b); the discussion of T numbers in the context of virus capsids²⁵ originates from this equation.

Plane-graph theory has been used to extend the original description of Goldberg polyhedra to create new families with more complicated

topologies and with more systematic notation; all possible topologies are tabulated in refs 26–28. We denote classical and extended Goldberg polyhedra by $\text{tri-G}(h, k)$ and $\text{tet-G}(h, k)$, respectively, where the prefixes ‘tri’ and ‘tet’ indicate trigonal and tetragonal nodes of the polyhedra. As indicated by plane-graph theory, the surfaces of $\text{tet-G}(h, k)$ polyhedra consist of eight triangles evenly distributed within a square grid network; the two indices h and k represent the relative positions of the closest triangles. The quadrangulation (Q) number is defined by the equation $Q = h^2 + k^2$, which describes the Pythagorean theorem ($Q = 1, 2, 4, 5, 8, 9, 10, 13, 16, \dots$) (Fig. 3c). For $Q = 1$, the notation $\text{tet-G}(1, 0)$ (which is equivalent to $\text{tet-G}(0, 1)$) indicates that the closest triangles are one step away from each other, thus giving rise to an octahedron (Fig. 3d). For $Q = 2$, $\text{tet-G}(1, 1)$ is specified, with the closest triangles located one step horizontally and one step vertically away from each other, resulting in a cuboctahedron (Fig. 3e). $Q = 4$ represents $\text{tet-G}(2, 0)$ (equivalent to $\text{tet-G}(0, 2)$), in which the closest triangles are two steps away from each other; this results in a rhombicuboctahedron (Fig. 3f).

Within this framework, our initially undefined complex **3** is a $\text{tet-G}(2, 1)$ polyhedron, with $Q = 5$. Both a schematic (Fig. 3g) and the crystallographic structure (Fig. 2a–c) reveal the closest triangles to be placed two steps horizontally and one step vertically away from each other. $Q = 5$ is the first Q number with an ‘asymmetric’ (h, k) combination: for $Q = 1, 2$ and 4 , the h and k indices are ‘symmetric’ in the sense that they can be exchanged without altering the corresponding structure (that is, $\text{tet-G}(h, k) = \text{tet-G}(k, h)$); however, for $Q = 5$, $\text{tet-G}(1, 2)$ and $\text{tet-G}(2, 1)$ give two mirror-image graphs and thus generate an enantiomeric pair of chiral frameworks (Fig. 3g). The graph presentation in Fig. 3d–g illustrates whether a given $\text{tet-G}(h, k)$ polyhedron is chiral or not (those for which $h \neq 0, k \neq 0$ and $h \neq k$ are chiral; all others are not). Using proven polyhedral formulae, primarily Euler’s polyhedron theorem, we determined the structures of the $\text{tet-G}(h, k)$ polyhedra for Q numbers up to 10 (Fig. 3h).

The underlying theory of the topology of M_nL_{2n} complexes (based on graph theory of Goldberg polyhedra) predicts that $\text{tet-G}(h, k)$ polyhedra with $Q \geq 8$ should exist. We deduced the structures of $M_{48}L_{96}$ ($\text{tet-G}(2, 2)$, $Q = 8$), $M_{54}L_{108}$ ($\text{tet-G}(3, 0)$, $Q = 9$) and $M_{60}L_{120}$ ($\text{tet-G}(3, 1)$, $Q = 10$) (Extended Data Fig. 2), and decided to target the predicted (but as yet unobserved) $M_{48}L_{96}$ complex (**4**)—the $\text{tet-G}(2, 2)$ polyhedron consisting of 8 triangles and 42 squares (Fig. 4a, inset).

We expected that $M_{30}L_{60}$ (**3**) would be a kinetically trapped metastable structure because modelling shows that the bend angle of 152° in ligand **1** should favour the less distorted $M_{48}L_{96}$. This result suggests that $M_{30}L_{60}$ could be converted to $M_{48}L_{96}$ under suitable conditions, prompting us to examine a range of self-assembly conditions (by changing the concentration, reaction time, temperature and so on) and the use of modified ligands. NMR spectra for the huge molecules in the $M_{48}L_{96}$ class of compounds are broadened, and so are not helpful for structural characterization. We were also not able to observe $M_{48}L_{96}$ or larger species using mass spectrometry, leaving X-ray crystallography as the only method capable of providing direct evidence for the assembly of $M_{48}L_{96}$ or larger structures. This X-ray crystallography involved carefully observing, under a microscope, the morphology of single crystals that formed directly upon extremely slow (over 2–3 months) vapour diffusion of isopropyl acetate into the dimethylformamide (DMF) reaction solutions. Diffraction data collected in preliminary screening experiments at a synchrotron facility indicated cell parameters close to those of **3** in most cases, but a very small fraction of the crystals taken from reaction solutions using forcing conditions (70°C for 48 h) showed a different diffraction pattern.

We collected data for more than 10 of the crystals with cell parameters different from those of **3**. Despite numerous attempts, a resolution of better than 2.85 \AA could not be obtained; the inherent properties of the compounds in the class we study—large dimensions and huge void spaces filled with solvents—make the compounds protein-like,

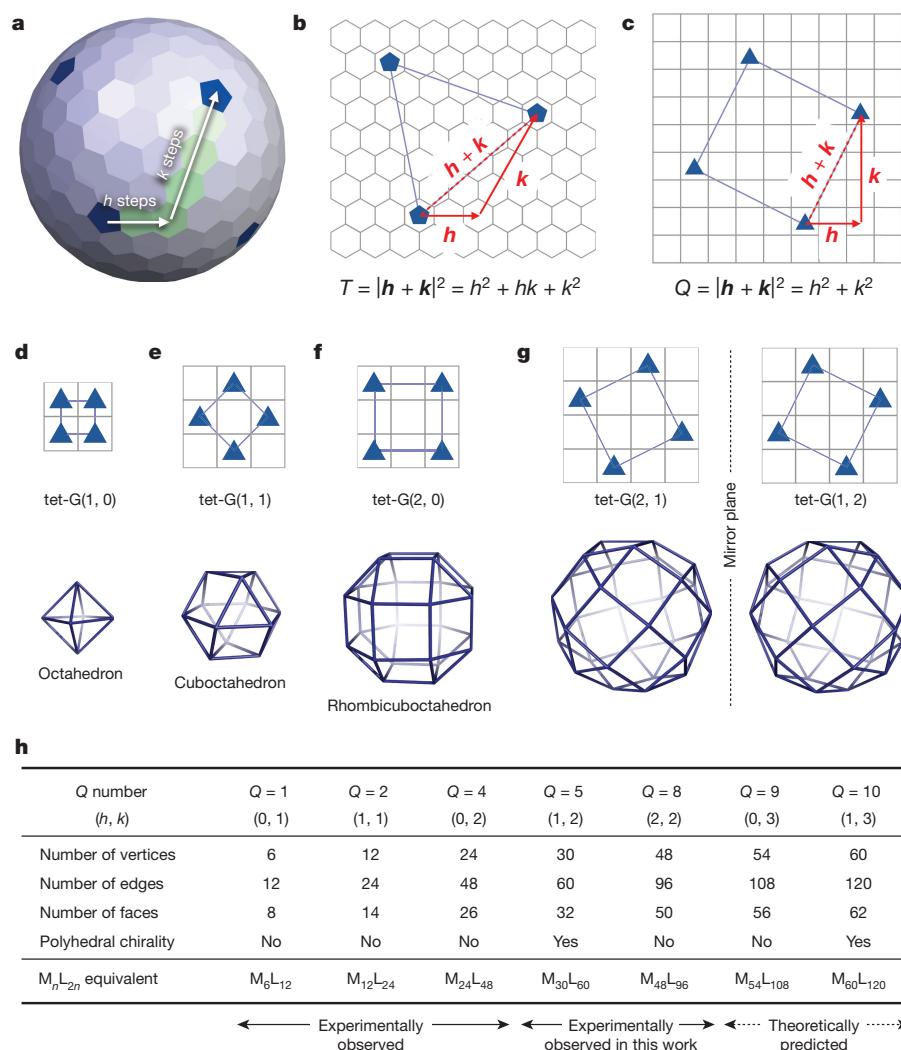


Figure 3 | Schematic representation of Goldberg polyhedra. **a**, Goldberg polyhedra consist of pentagons and hexagons. Each polyhedron is denoted by the locational relationship between the closest pair of pentagons. In this case, the polyhedron is denoted tri-G(2, 4) ($h=2$, $k=4$). **b**, Definition of *T* number. The *T* number represents the square of the distance between the two closest pentagons on the hexagonal network. From the Pythagorean theorem, $T = |\mathbf{h} + \mathbf{k}|^2 = [|\mathbf{h}| + |\mathbf{k}|\cos(60^\circ)]^2 + [|\mathbf{k}|\sin(60^\circ)]^2 = h^2 + hk + k^2$. **c**, Definition of *Q* number. By simple analogy with the *T* number, the *Q* number is defined by the square of the distance between the two closest triangles on the square grid network. **d–f**, Diagrams

of tet-G(*h*, *k*) polyhedra. tet-G(1, 0) with $Q = 1$ is equivalent to an octahedron (**d**), tet-G(1, 1) with $Q = 2$ is a cuboctahedron (**e**) and tet-G(2, 0) with $Q = 4$ a rhombicuboctahedron (**f**). **g**, Diagram of tet-G(2, 1) and tet-G(1, 2) ($Q = 5$), which have topology that is identical to that of the left- and right-hand forms of $M_{30}L_{-60}$, respectively, shown in Fig. 1a, b. The 'asymmetric' *h* and *k* values support the origin of chirality. **h**, Summary of extended Goldberg polyhedra ordered by *Q* number. Only those corresponding to $Q = 1, 2$ and 4 have been observed experimentally previously; in this work we observe those corresponding to $Q = 5$ and 8 ; and the others are predicted by theory, but have yet to be observed.

so we cannot expect to achieve the higher resolution that is typical for standard metal complexes studied using conventional molecular crystallography. Several of the best diffraction datasets were subjected to structure refinement.

The obtained electron density map clearly shows a very large spherical entity that consists of a framework made up of triangles and squares. Although the large molecular diameter and the disordered solvent molecules within the huge void space (70% of the unit cell volume) substantially attenuated the diffraction intensity, particularly in the high-angle region, strong peaks from heavy atoms (palladium atoms at the nodes and selenium atoms in the middle of the edges) made the structure refinement possible (Fig. 4b, c). The positions of the selenium and palladium atoms were clearly refined, and then ligand **1**, which bridges the palladium centres, was modelled with its density function theory (DFT)-optimized structure, as is common in protein crystallography (Fig. 4b). From the refined structure, we were able to confirm the overall geometry of the tet-G(2, 2) polyhedron ($Q = 8$) and, hence, the successful self-assembly of the $M_{48}L_{96}$ complex. Within the

large spherical framework, the closest triangles are in an (h , k) = (2, 2) relationship—two steps horizontally and two steps vertically away from each other (Fig. 4d). Our crystallographic data provide a clear visualization of this tet-G(2, 2) polyhedral framework.

Although it was a serendipitous experimental observation that led us to explore the theory of tetravalent Goldberg polyhedra, this theory then enabled us to consider and achieve the self-assembly of the giant $M_{48}L_{96}$ polyhedron, which contains the largest number of components observed in a self-assembled molecular structure so far. Tetravalent Goldberg polyhedra, which have square packing, appear rarely in nature; by contrast, hexagonal packing—as seen in graphite and honeycombs and in trivalent Goldberg polyhedra such as fullerenes and virus capsids—is frequently observed. This reason for this difference in prevalence is presumably that square packing is not as close as hexagonal packing. The square planar coordination geometry of palladium(II) ions is a rare example of a square motif that can induce the unusual square packing and, here, that enabled the formation of tetravalent Goldberg polyhedra. Given that quasi-crystals were once

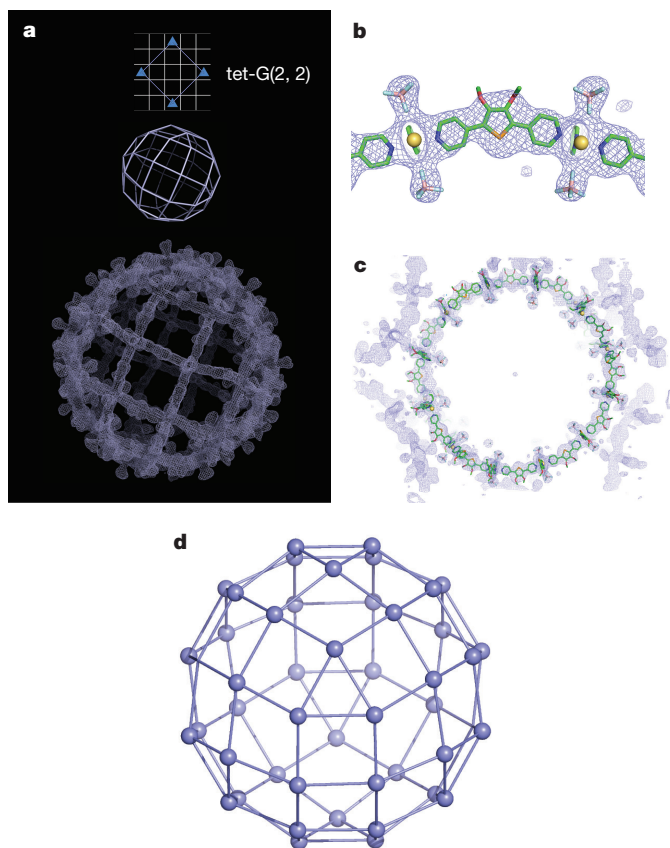


Figure 4 | X-ray crystallographic analysis of self-assembled product 4. **a**, Entire view of the MEM electron density map (0.6 electrons per \AA^3) for $M_{48}L_{96}$ complex **4** (right). This complex has the topology of the tet-G(2, 2) polyhedron with $Q = 8$ (left). Supplementary Video 2 shows a rotating electron density map. **b**, Enlarged view around a ligand in $M_{48}L_{96}$ with its MEM electron density map. The modelled ligand structure agrees well with the observed electron density. Space group, $I\bar{4}$; lattice parameters, $a = b = 63.7 \text{ \AA}$, $c = 94.6 \text{ \AA}$; data resolution, 2.85 \AA ; $R(I > 2\sigma(I))$, 0.2181; $R_{\text{free}}(I > 2\sigma(I))$, 0.2465. **c**, Sliced image of the $M_{48}L_{96}$ complex with the MEM electron density map. No distinguishable peaks in electron density are observed in the void space; this supports the validity of the model building. **d**, Crystal structure of **4** emphasizing the tet-G(2, 2) topology and metal centres.

considered obscure, but are now rationally designed²⁹ and known to be ubiquitous in many areas³⁰, we expect that many other examples of tetravalent Goldberg polyhedral will be discovered or artificially manufactured in the near future.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 August; accepted 25 October 2016.

- Conn, M. M. & Rebek, J. Self-assembling capsules. *Chem. Rev.* **97**, 1647–1668 (1997).
- Caulder, D. L. & Raymond, K. N. Supramolecules by design. *Acc. Chem. Res.* **32**, 975–982 (1999).
- Saalfraank, R. W., Maid, H. & Scheurer, A. Supramolecular coordination chemistry: the synergistic effect of serendipity and rational design. *Angew. Chem. Int. Ed.* **47**, 8794–8824 (2008).
- Forgan, R. S., Sauvage, J.-P. & Stoddart, J. F. Chemical topology: complex molecular knots, links, and entanglements. *Chem. Rev.* **111**, 5434–5464 (2011).
- Chakrabarty, R., Mukherjee, P. S. & Stang, P. J. Supramolecular coordination: self-assembly of finite two- and three-dimensional ensembles. *Chem. Rev.* **111**, 6810–6918 (2011).

- Young, N. J. & Hay, B. P. Structural design principles for self-assembled coordination polygons and polyhedra. *Chem. Commun.* **49**, 1354–1379 (2013).
- Smulders, M. M. J., Riddell, I. A., Browne, C. & Nitschke, J. R. Building on architectural principles for three-dimensional metallosupramolecular construction. *Chem. Soc. Rev.* **42**, 1728–1754 (2013).
- Cook, T. R. & Stang, P. J. Recent developments in the preparation and chemistry of metallacycles and metallacages via coordination. *Chem. Rev.* **115**, 7001–7045 (2015).
- Hsia, Y. *et al.* Design of a hyperstable 60-subunit protein icosahedron. *Nature* **535**, 136–139 (2016).
- Zlotnick, A. To build a virus capsid: an equilibrium model of the self assembly of polyhedral protein complexes. *J. Mol. Biol.* **241**, 59–67 (1994).
- Harris, K., Fujita, D. & Fujita, M. Giant hollow M_nL_{2n} spherical complexes: structure, functionalisation and applications. *Chem. Commun.* **49**, 6703–6712 (2013).
- Fujita, D., Yokoyama, H., Ueda, Y., Sato, S. & Fujita, M. Geometrically restricted intermediates in the self-assembly of an $M_{12}L_{24}$ cuboctahedral complex. *Angew. Chem. Int. Ed.* **54**, 155–158 (2015).
- Suzuki, K., Tominaga, M., Kawano, M. & Fujita, M. Self-assembly of an M_6L_{12} coordination cube. *Chem. Commun.* 1638–1640 (2009).
- Tominaga, M. *et al.* Finite, spherical coordination networks that self-organize from 36 small components. *Angew. Chem. Int. Ed.* **43**, 5621–5625 (2004).
- Sun, Q. F. *et al.* Self-assembled $M_{24}L_{48}$ polyhedra and their sharp structural switch upon subtle ligand variation. *Science* **328**, 1144–1147 (2010).
- Fujita, D. *et al.* Self-assembly of $M_{30}L_{60}$ icosidodecahedron. *Chem* **1**, 91–101 (2016).
- Coxeter, H. S. M. *Introduction to Geometry* 2nd edn, Ch. 10, 21 (Wiley, 1989).
- Goldberg, M. A class of multi-symmetric polyhedra. *Tohoku Math. J.* **43**, 104–108 (1937).
- Schein, S. & Gayed, J. M. Fourth class of convex equilateral polyhedron with polyhedral symmetry related to fullerenes and viruses. *Proc. Natl Acad. Sci. USA* **111**, 2920–2925 (2014).
- Caspar, D. L. D. & Klug, A. Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24 (1962).
- Fowler, P. W. & Manolopoulos, D. E. *An Atlas of Fullerenes* Ch. 2 (Courier Corporation, 2006).
- Bunzen, M. *et al.* Self-assembly of $M_{24}L_{48}$ polyhedra based on empirical prediction. *Angew. Chem. Int. Ed.* **51**, 3161–3163 (2012).
- Takata, M. The MEM/Rietveld method with nano-applications — accurate charge-density studies of nano-structured materials by synchrotron-radiation powder diffraction. *Acta Crystallogr. A* **64**, 232–245 (2008).
- Fujita, D. *et al.* Protein encapsulation within synthetic molecular hosts. *Nat. Commun.* **3**, 1093 (2012).
- Prasad, B. V. & Schmid, M. F. Principles of virus structural organization. *Adv. Exp. Med. Biol.* **726**, 17–47 (2012).
- Sikirić, M. D. & Deza, M. in *The Mathematics and Topology of Fullerenes* (eds Cataldo, F. *et al.*) Ch. 6, 103–116 (Springer, 2011).
- Dutour, M. & Deza, M. Goldberg–Coxeter construction for 3- and 4-valent plane graphs. *Electron. J. Combin.* **11**, R20 (2004).
- Brinkmann, G. & Deza, M. Lists of face-regular polyhedra. *J. Chem. Inf. Comput. Sci.* **40**, 530–541 (2000).
- Zhou, Z. & Harris, K. D. M. Design of a molecular quasicrystal. *ChemPhysChem* **7**, 1649–1653 (2006).
- Vekilov, Y. K. & Chernikov, M. A. Quasicrystals. *Phys. Uspekhi* **53**, 537–560 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the JST–PRESTO programme and the ACCEL programme, and partly supported by KAKENHI (25102007). The synchrotron X-ray crystallography was performed at the BL38B1 and BL41XU beamlines at SPring-8 (2014A0042 and 2015B0120); preliminarily experiments were performed at the BL-1A beamline at KEK PF (2015G097). We thank M. Kotani and M. Tagami for providing us with mathematical discussion (see Methods) and references (refs 26–28).

Author Contributions D.F. and Y.U. performed and analysed the experiments. D.F. and S.S. worked on the preliminary X-ray diffraction analysis. N.M. and T.K. finalized the X-ray data. D.F. built the mathematical discussion. D.F. and M.F. wrote the paper. D.F. and M.F. designed and supervised the research project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.F. (fujitadaiishi@appchem.t.u-tokyo.ac.jp) and M.F. (mfujita@appchem.t.u-tokyo.ac.jp).

Reviewer Information *Nature* thanks F. Beuerle, P. Schwerdtfeger and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

General information. NMR spectra were measured on a Bruker DRX-500 (500 MHz) spectrometer equipped with a BBO gradient probe and a Bruker AV-500 (500 MHz) spectrometer equipped with a TCI gradient CryoProbe. Infrared spectra were measured as KBr pellets using a DIGILAB FTS-7000 instrument. Melting points were determined on a Yanaco MP-500V apparatus. Solvents and reagents were purchased from TCI, WAKO Pure Chemical Industries and Sigma-Aldrich. Unless otherwise noted, all of the chemicals were reagent grade and used without any further purification.

Synthesis of ligand 1. 4-pyridylboronic acid pinacol ester (440 mg, 2.2 mmol), PdCl₂(dppf) (50 mg, 0.072 mmol) and K₃PO₄ (910 mg, 4.3 mmol) were added at room temperature to a solution of 2,5-dibromo-3,4-dimethoxyselenophene (270 mg, 0.72 mmol) in toluene/*i*-PrOH/H₂O (5.0 ml, 3/1/1), and the resulting mixture was stirred at 85 °C for 16 h. The reaction mixture was filtered through a celite pad and washed with AcOEt. The filtrate was washed with water and brine, dried over anhydrous sodium sulfate and evaporated *in vacuo*. The residue was purified by column chromatography on silica gel (chloroform:methanol = 100:0 → 98:2) to give **1** as a brown solid. The resultant material was purified by gel permeation chromatography (eluent: CHCl₃) to afford a pale yellow solid (130 mg, 53%). Melting point: 120–121 °C. ¹H NMR (500 MHz, CDCl₃, 27 °C) δ (p.p.m.): 8.62 (d, *J* = 6.0 Hz, 4H), 7.57 (d, *J* = 6.0 Hz, 4H), 3.91 (s, 6H). ¹³C NMR (125 MHz, CDCl₃, 27 °C) δ (p.p.m.): 150.7 (C), 150.4 (CH), 141.1 (C), 127.2 (C), 121.1 (CH), 60.3 (CH₃). Infrared (cm^{−1}): 3,206, 3,040, 2,846, 1,589, 1,540, 1,489, 1,415, 1,370, 1,321, 1,292, 1,224, 1,214, 1,114, 1,099, 1,070, 1,030, 1,000, 990, 900, 864, 818. Mass spectroscopy (MS) (EI): calculated for *m/z* C₁₆H₁₄O₂N₂Se [M⁺] 346.0, found 346.0. HRMS (ESI TOF): calculated for *m/z* C₁₆H₁₅O₂N₂Se [M + H]⁺ 347.0294, found 347.0297. Elemental analysis: calculated for C₁₆H₁₄O₂N₂Se: C, 55.66; H, 4.09; N, 8.11. Found: C, 55.52; H, 4.12; N, 8.07. See Extended Data Figs 3 and 4 for full range NMR spectra.

Self-assembly of M₃₀L₆₀ complex 3. A solution of [Pd(BF₄)₂](CH₃CN)₄ in DMSO-*d*₆ (5.2 μmol) was added to a solution of ligand **1** (3.14 mg, 10 μmol) in DMSO-*d*₆ (0.50 ml) and the resulting mixture was stirred at room temperature for 3 h. The reaction can also be carried out in DMF-*d*₇. ¹H NMR (500 MHz, DMSO-*d*₆, 300 K) δ (p.p.m.): 9.2 (br), 7.9 (br), 3.8 (br). ¹H NMR (500 MHz, DMF-*d*₇, 300 K) δ (p.p.m.): 9.4 (br), 8.0 (br), 3.9 (br).

The ¹H nuclear magnetic resonance (¹H NMR) spectrum showed considerable broadening and downfield shifts of the peaks assigned to the aromatic protons (Extended Data Fig. 1b). Downfield shifts of PyH_α (Δδ = 0.57 p.p.m.) and PyH_β (Δδ = 0.22 p.p.m.) suggested that the pyridines were coordinated to palladium(II) ions, and the broadened signals were consistent with the formation of a very large molecular structure with high symmetry. ¹H diffusion-ordered NMR spectroscopy (¹H DOSY) also indicated the formation of a single product with a diffusion coefficient *D* of 2.6 × 10^{−11} m² s^{−1} (log *D* = −10.58) at 300 K in DMSO-*d*₆ (Extended Data Fig. 1c). The observed value of *D* is smaller than that for the corresponding M₂₄L₄₈ complex⁴, which suggests the formation of a larger product. See Extended Data Figs 5–8 for full-range NMR spectra. Reliable evidence for the formation of **3** was obtained by the crystallographic analysis described below.

Self-assembly of M₄₈L₉₆ complex 4. M₄₈L₉₆ complex **4** was obtained as a side product of **3** when the complex was prepared under forcing conditions (70 °C for 48 h). Presumably, complex **3** is a kinetically trapped product, which in part turned into thermodynamic product **4** at 70 °C. The ratio **3**:**4** could not be determined because of the complete overlap of broadened signals of the two complexes in ¹H NMR. They were not distinguished by DOSY. From this solution, single crystals of **3** and **4** were directly obtained as a mixture by extremely slow (over 2–3 months) vapour diffusion of isopropyl acetate into the DMF reaction solutions. Under a microscope, we tried to distinguish the single crystals of **3** and **4** from the difference

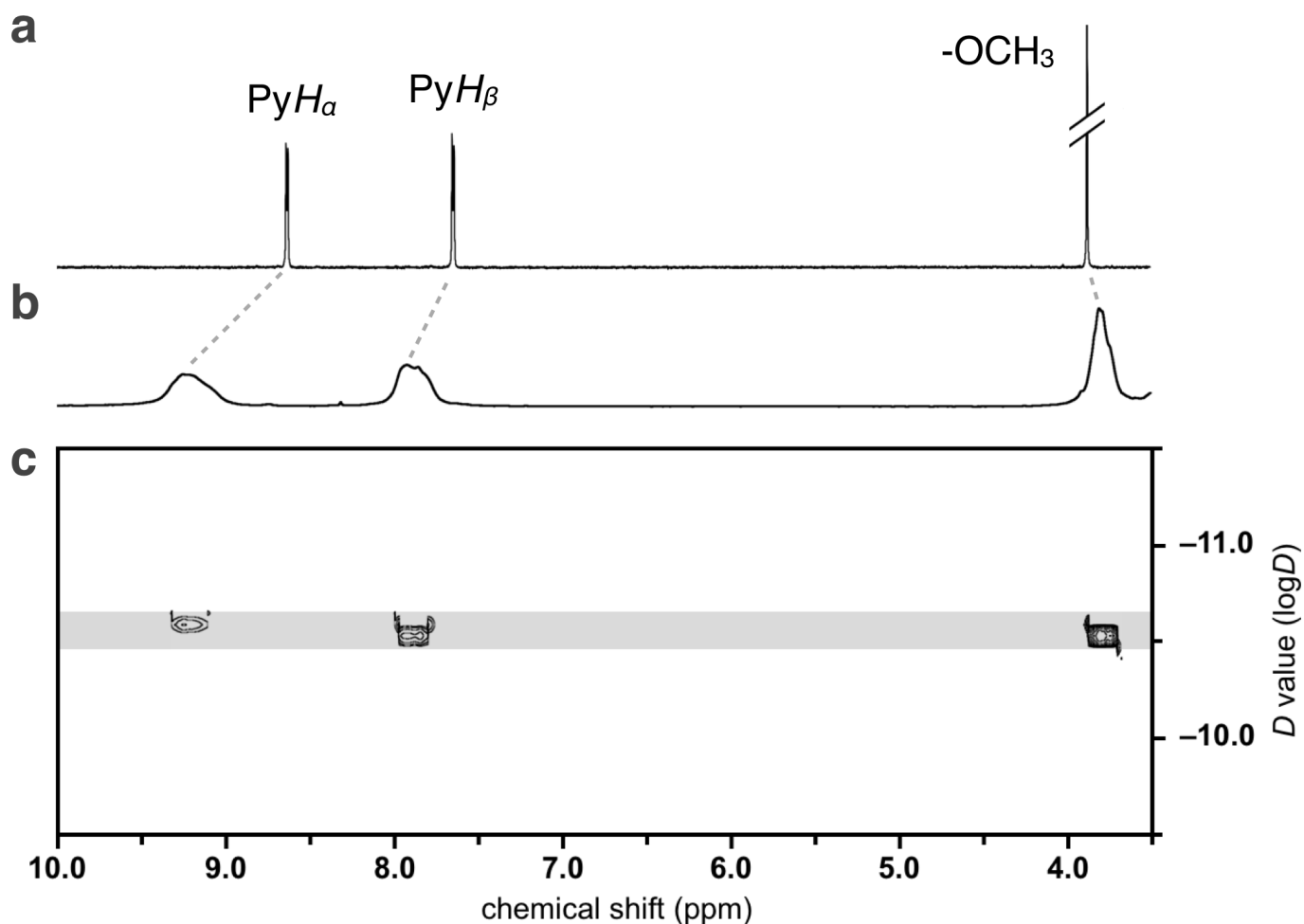
in the crystal morphology. Although most of the crystals were diffracted with the cell parameters of **3**, a very small portion of the crystals were diffracted in a different pattern in preliminary screening experiments at a synchrotron facility, and were subsequently characterized as **4** by the crystallographic analysis.

Crystallographic analysis of 3 and 4. Single crystals of **3** were prepared under extremely slow vapour diffusion (over 2–3 months) of isopropyl acetate into an *N,N*-DMF solution of **3** (with tetrafluoroborate as the counter anion). Single crystals of **4** were obtained in a similar way. The final diffraction data were measured at the BL38B1 beamline (Rayonix MX225HE detector) and the BL41XU beamline (PILATUS 6M detector) at SPring-8. XDS³¹ was used for the processing and data reduction of **3** and **4**. The structures were solved and refined by SHELXT³² and SHELXL³³. These data were not of high enough resolution to allow calculation of the hydrogen positions. The crystal lattice contains huge voids (75.1% for **3** and 70.2% for **4**) with a large number of disordered solvents and anions. Owing to the very poor diffraction power of the crystal (resolution of 1.95 Å for **3** and 2.85 Å for **4**), solvent molecules could not be located as realistic molecular structures even in the maximum entropy density map calculated by ENIGMA³⁴. The ligands showed very large thermal motion (maybe as a result of positional disorder, which, however, could not be modelled) and so many restraints on the bond distances and thermal parameters (DFIX, DANG and FLAT) had to be applied to the ligands, counter anions and Pd–N distances to avoid chemically unreasonable bond distances and angles. Pd atoms were refined anisotropically; all of the other atoms were refined isotropically. ENIGMA was used to calculate a reasonable electron density map with the MEM. For **3**, the total number of electrons *T_E* ≈ 321,000 of the crystal is determined by summing the modelled electrons of M₃₀L₆₀ and BF₄ (14,280 × 8) and product of the volume of solvent (approximately 581,500 Å³) and the average electron density of DMSO (0.356 electrons per Å³). The DMSO solvent molecules were determined from the MEM map. For **4**, *T_E* ≈ 142,000 is determined as the sum of the modelled electrons of M₃₀L₆₀ and BF₄ (45,696) and the product of the volume of solvent (approximately 269,438 Å³) and the average electron density of DMSO (0.356 electrons per Å³). The crystallographic data are summarized in Extended Data Table 1.

Mathematical discussion. The surface of any convex polyhedron has Euler characteristic *V* − *E* + *F* = 2, where *V*, *E* and *F* are the numbers of vertices, edges and faces, respectively, of the given polyhedron. In addition, the following equation for tetravalent Goldberg polyhedra is derived from equation (53) in ref. 35: 3*F*₃ + 2*F*₄ − 2*V*₄ = 12, where *F*₃ is the number of triangles, *F*₄ is the number of squares and *V*₄ is the number of 4-valent vertices. If we set the number of triangles to 8, then we get a very simple formula for the tet-*G*(*h*, *k*) structures discussed here: *F* − *V* = 2 (for tetravalent Goldberg polyhedra, *F* = *F*₃ + *F*₄ and *V* = *V*₄). Because *V* = 6*Q*, the values of *V*, *E* and *F* can be determined for any tet-*G*(*h*, *k*) polyhedra.

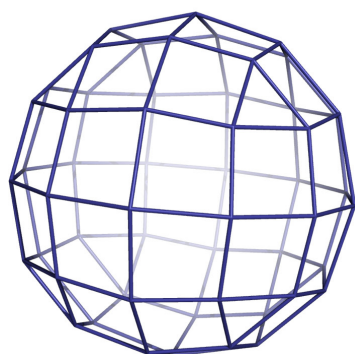
Data availability. All of the data generated and analysed during this study are included in this Letter, its Extended Data and its Supplementary Information.

- Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
- Sheldrick, G. M. SHELXT — integrated space-group and crystal-structure determination. *Acta Crystallogr. A* **71**, 3–8 (2015).
- Sheldrick, G. M. Crystal structure refinement with SHELXL. *Acta Crystallogr. C* **71**, 3–8 (2015).
- Tanaka, H. *et al.* ENIGMA: maximum-entropy method program package for huge systems. *J. Appl. Cryst.* **35**, 282–286 (2002).
- Schwerdtfeger, P., Wirz, L. N. & Avery, J. The topology of fullerenes. *WIREs Comp. Mol. Sci.* **5**, 96–145 (2015).
- Yokoyama, H., Ueda, Y., Fujita, D., Sato, S. & Fujita, M. Finely resolved threshold for the Sharp M₁₂L₂₄/M₂₄L₄₈ structural switch in multi-component M_nL_{2n} polyhedral assemblies: X-ray, MS, NMR, and ultracentrifugation analyses. *Chem. Asian J.* **10**, 2292–2295 (2015).

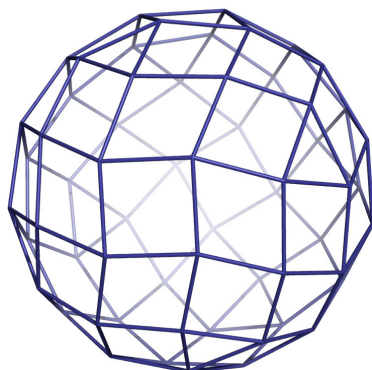


Extended Data Figure 1 | NMR study of the self-assembly of ligand 1.
a, ^1H NMR spectrum (500 MHz, $\text{DMSO}-d_6$, 300 K) of ligand 1. The signal denoted PyH_α is derived from the protons in the pyridyl α -position; that denoted PyH_β is derived from the protons in the pyridyl β -position. The signal denoted $-\text{OCH}_3$ is from the methoxy protons. **b**, ^1H NMR spectrum of 1 after self-assembly with palladium(II) ions (BF_4^- salt).

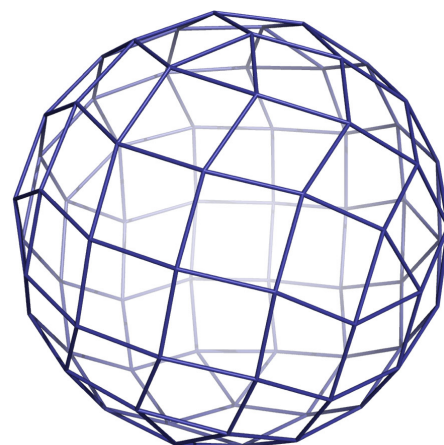
The aromatic signals are shifted downfield and heavily broadened. **c**, ^1H DOSY spectrum of 1 after self-assembly with palladium(II) ions (BF_4^- salt). The spectrum indicates a single product with a diffusion coefficient D of $2.6 \times 10^{-11} \text{ m}^2 \text{ s}^{-1}$ ($\log D = -10.58$). The grey band is a guide to the eye. All of the NMR spectra (500 MHz) were measured for $\text{DMSO}-d_6$ solutions at 300 K.



tet-G(0,3)
 $Q = 9$

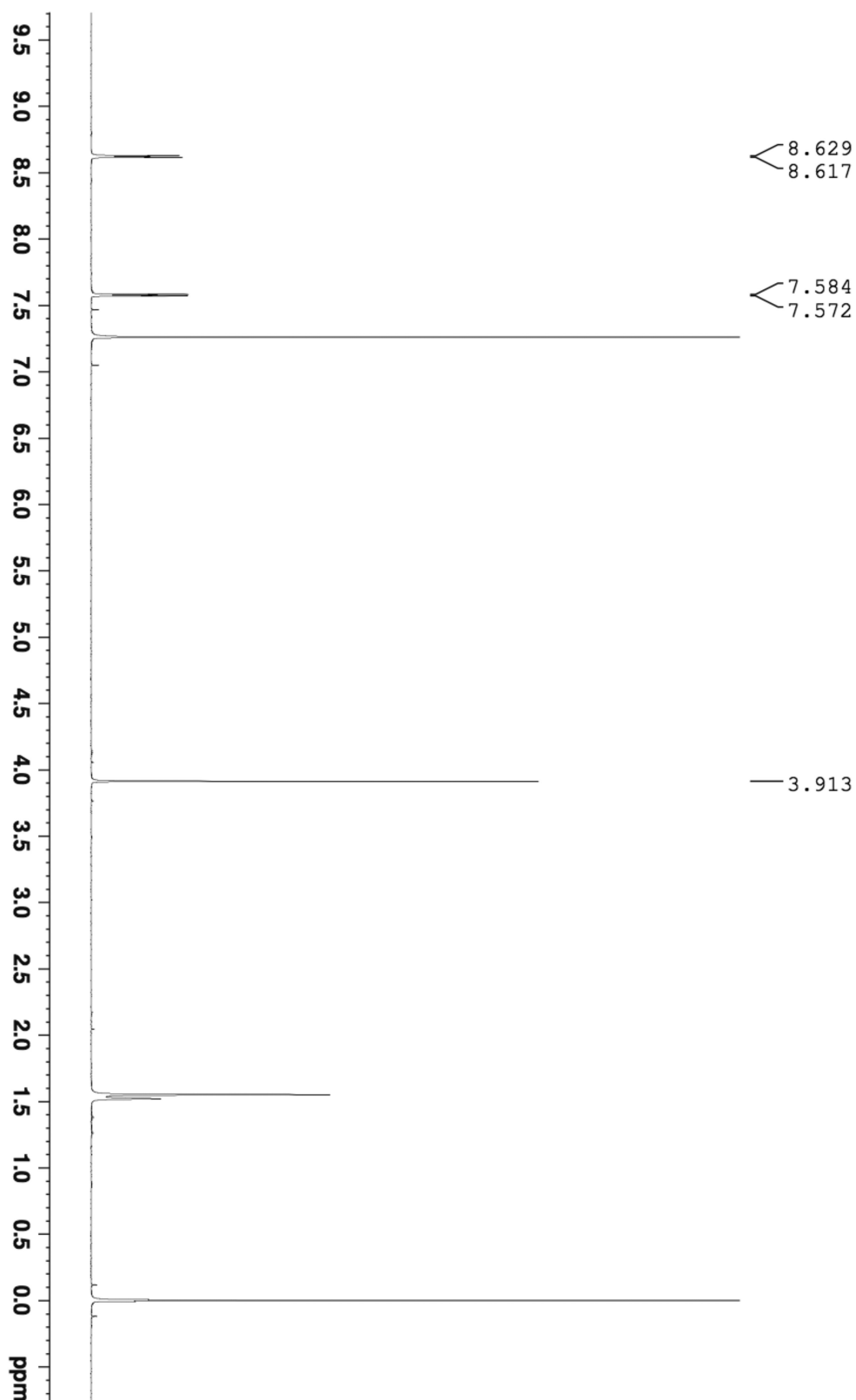


tet-G(1,3)
 $Q = 10$

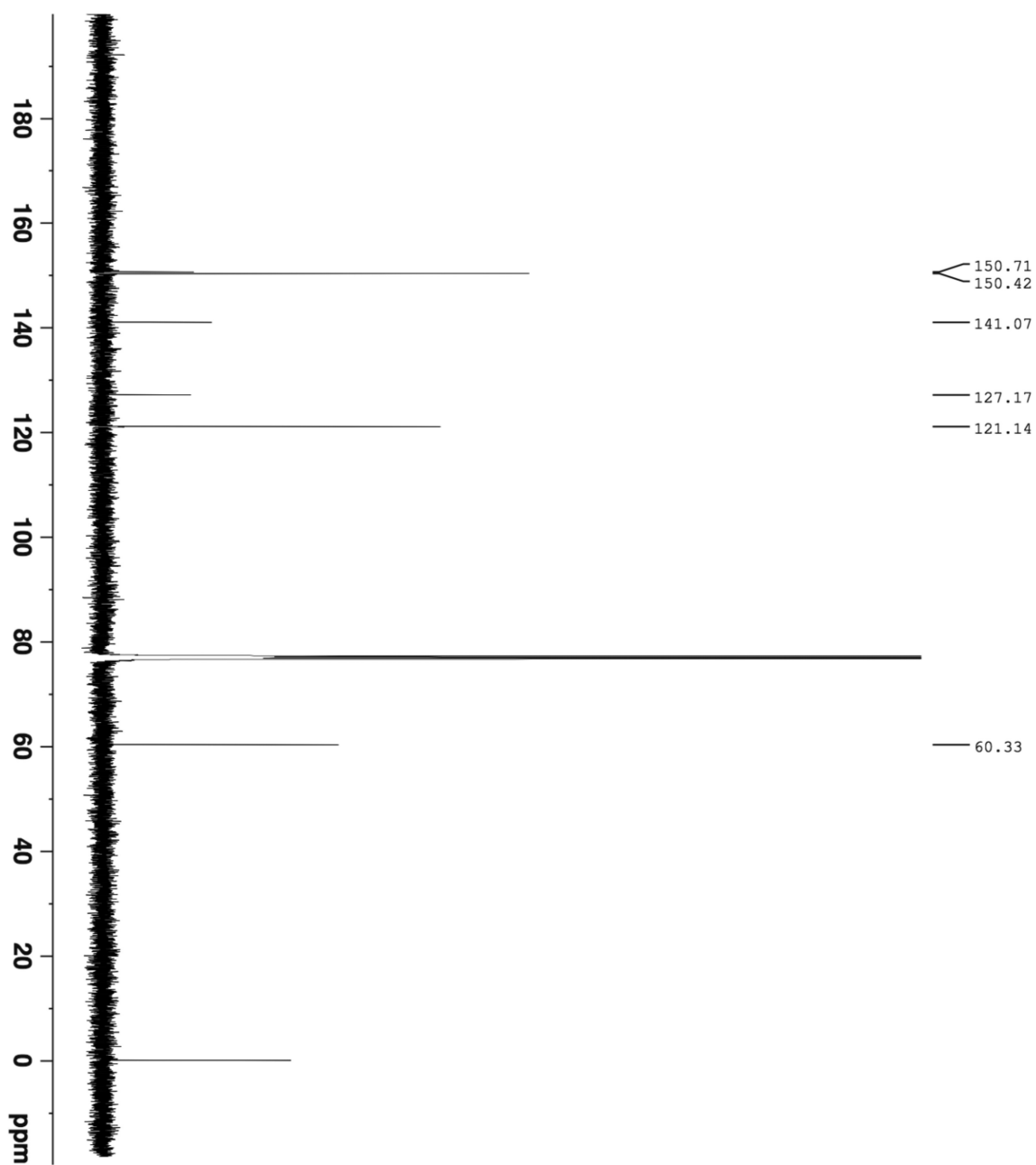


tet-G(2,3)
 $Q = 13$

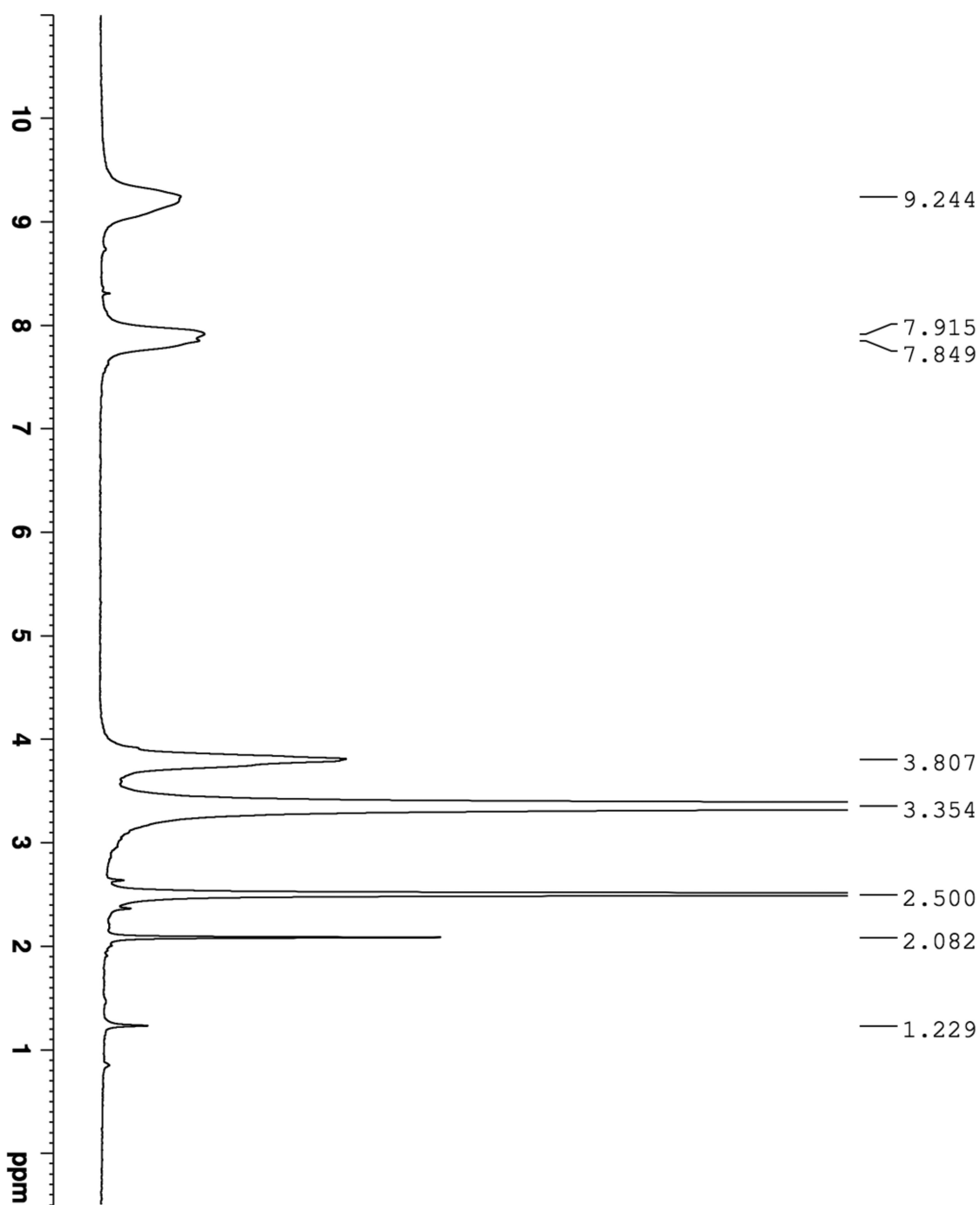
Extended Data Figure 2 | Schematic representation of larger *tet-G(h, k)* polyhedra. Polyhedra with the topology of *tet-G(3, 0)* (or, equivalently of *tet-G(0, 3)*; $Q = 9$), *tet-G(1, 3)* ($Q = 10$) or *tet-G(2, 3)* ($Q = 13$). For $Q = 9$ and $Q = 10$, the other structure in the chiral pair (*tet-G(3, 1)* and *tet-G(3, 2)*, respectively) is a mirror image of the polyhedron shown.



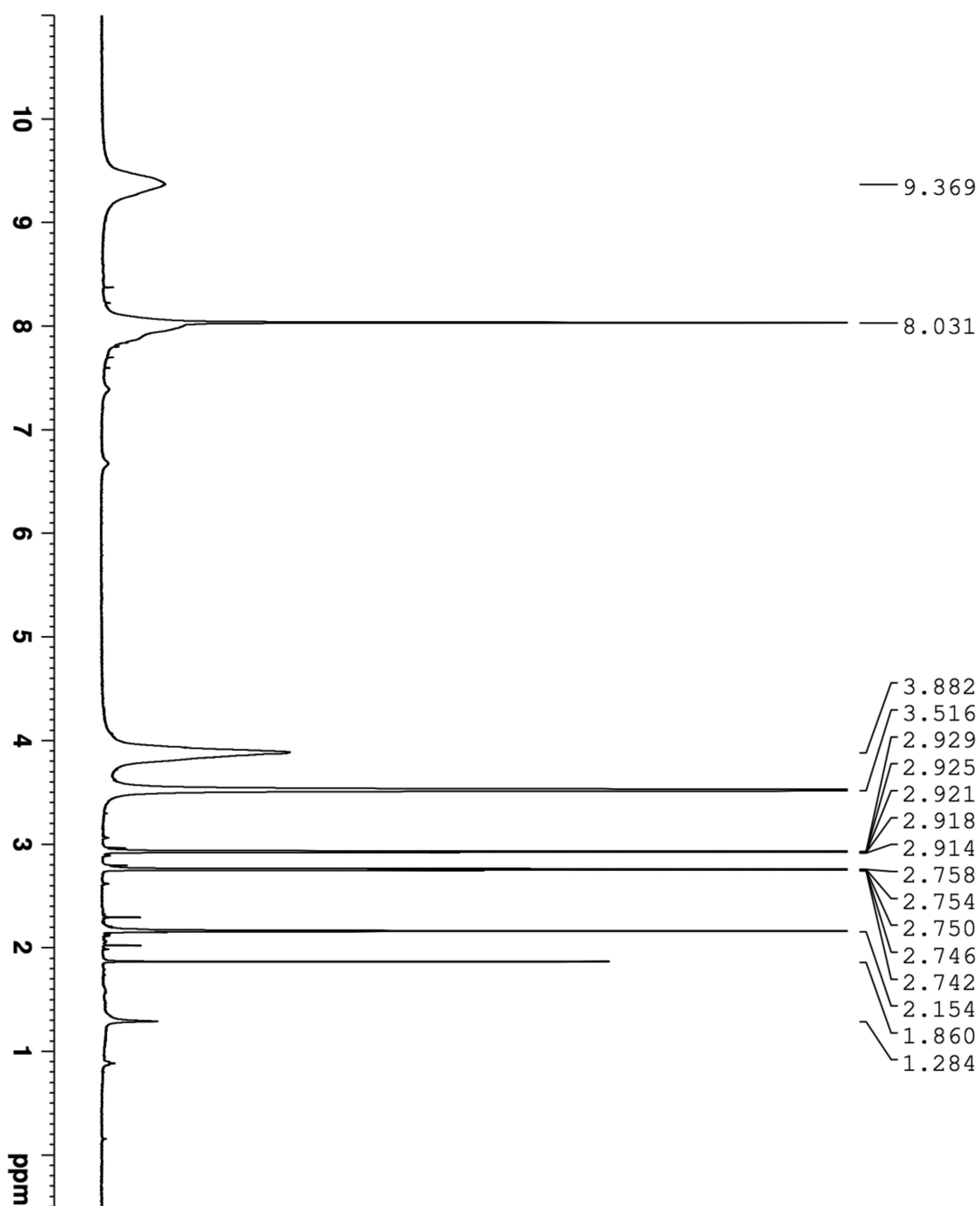
Extended Data Figure 3 | ¹H NMR of ligand 1. Full range spectrum: 500 MHz, CDCl₃, 27°C.



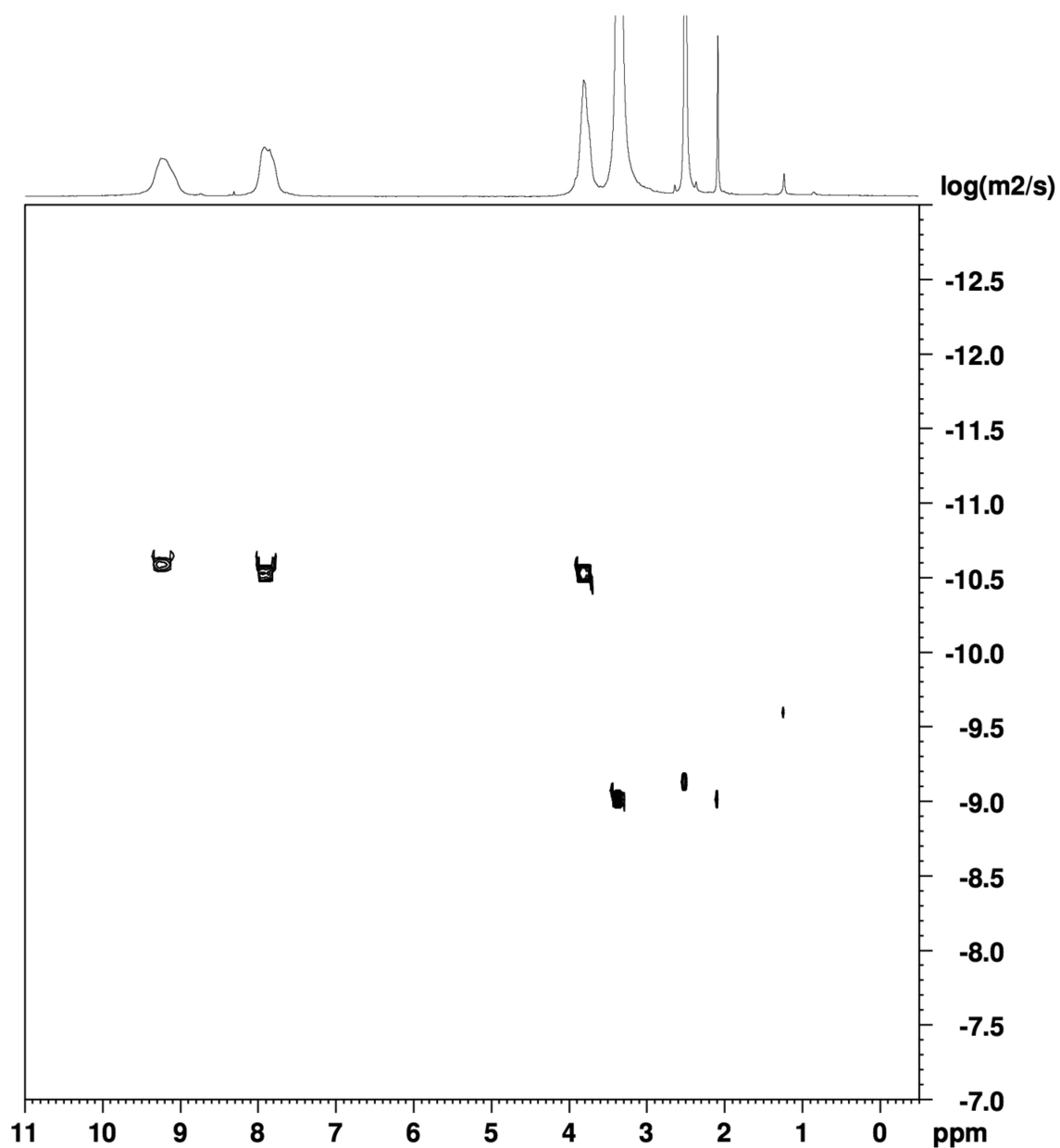
Extended Data Figure 4 | ¹³C NMR of ligand 1. Full range spectrum: 125 MHz, CDCl₃, 27 °C.



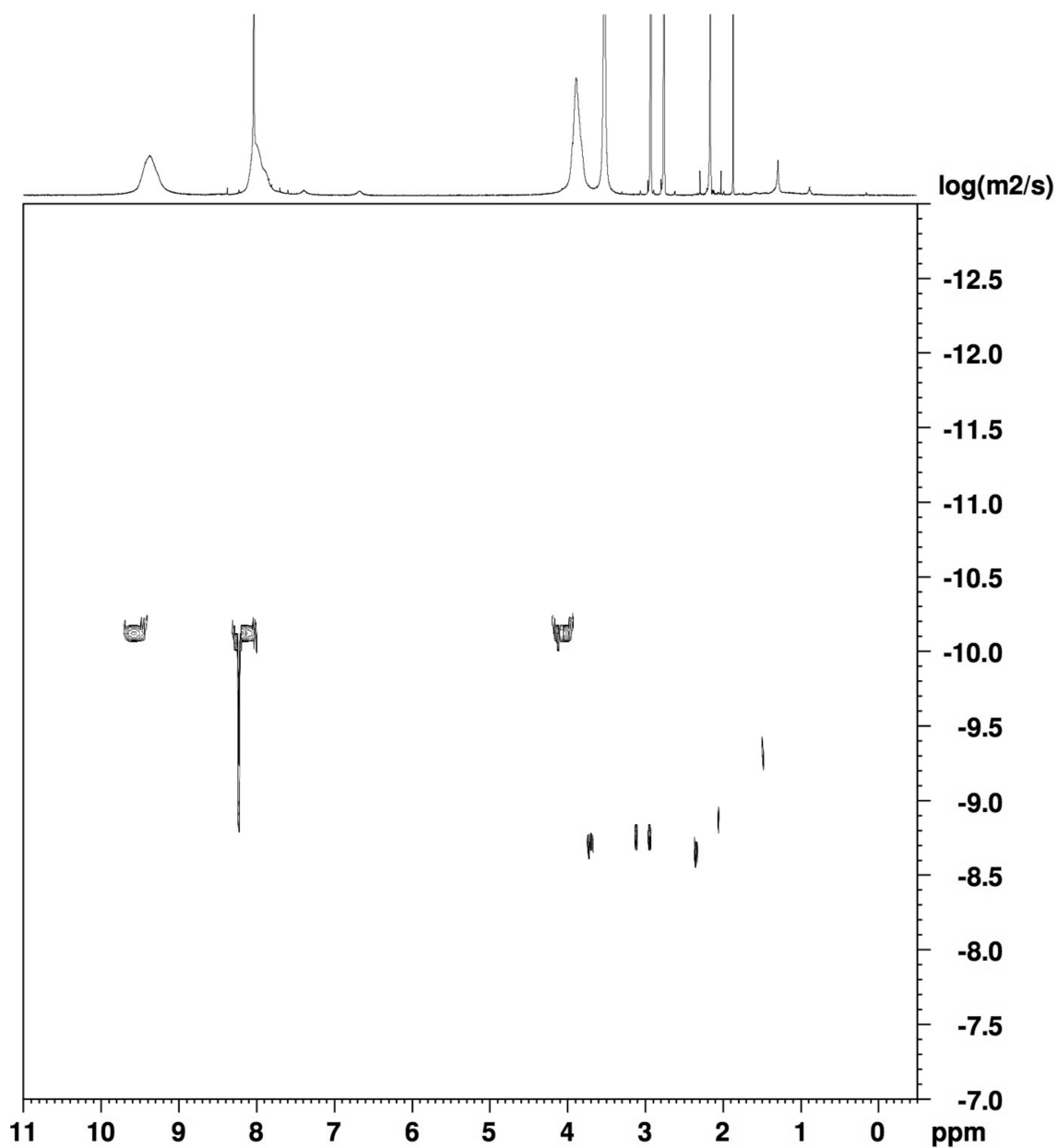
Extended Data Figure 5 | ¹H NMR of 3. Full range spectrum: 500 MHz, DMSO-*d*₆, 27 °C.



Extended Data Figure 6 | ¹H NMR of 3. Full range spectrum: 500 MHz, DMF-*d*₅, 27 °C.



Extended Data Figure 7 | ^1H DOSY NMR of **3**. Full range spectrum: 500 MHz, $\text{DMSO}-d_6$, 27 °C. For comparison with $\text{M}_{12}\text{L}_{24}$ or $\text{M}_{24}\text{L}_{48}$ complexes, see ref. 36.



Extended Data Figure 8 | ^1H DOSY NMR of 3. Full range spectrum: 500 MHz, $\text{DMF}-d_5$, 27°C. For comparison with $\text{M}_{12}\text{L}_{24}$ or $\text{M}_{24}\text{L}_{48}$ complexes, see ref. 36.

Extended Data Table 1 | Crystal data and structural refinement for M₃₀L₆₀ (3) and M₄₈L₉₆ (4)

	M ₃₀ L ₆₀ (3)	M ₄₈ L ₉₆ (4)
Measurement		
X-ray Source	SPring-8 BL38B1	SPring-8 BL41XU
Detector	Rayonix MX225HE	PILATUS 6M
Temperature	100 K	296(2) K
Wavelength	0.97900 Å	0.70000 Å
Crystal system	Centric	Centric
Crystal size	100 × 100 × 80 mm ³	150 × 150 × 50 mm ³
Data processing		
Space group	<i>Pbca</i>	<i>I</i> $\bar{4}$
Unit cell dimensions	a = 73.0 (1) Å	a = b = 63.7 (3) Å
	b = 73.2 (1) Å	c = 94.6 (3) Å
	c = 143.4 (2) Å	
Volume	774959.56 Å ³	383816.78 Å ³
Resolution	100.0 – 1.95 (2.06 – 1.95) Å	100.0 – 2.85 (3.02 – 2.85) Å
Absorption correction	Empirical	Empirical
Reflections collected	403060	30333
Independent reflections	54165	8637
Completeness	99.6 (97.4) %	99.6 (98.8) %
R _{merge}	0.160 (0.887)	0.108 (0.682)
I/σ	8.52 (2.35)	7.87 (1.64)
Redundancy	7.44 (7.36)	3.51 (3.19)
Refinement		
Empirical formula	C ₇₆₈₀ H ₆₇₂₀ N ₉₆₀ O ₉₆₀	C ₃₀₇₂ H ₂₆₈₈ N ₃₈₄ O ₃₈₄
	Se ₄₈₀ Pd ₂₄₀ B ₄₈₀ F ₁₉₂₀	Se ₁₉₂ Pd ₉₆ B ₁₉₂ F ₇₆₈
Data / restraints / parameters	48312 / 4440 / 6813	8234 / 1818 / 2543
Goodness-of-fit on F ²	3.309	3.725
R (I > 2σ(I))	0.2048	0.2198
R _{free} (I > 2σ(I))	0.2404	0.2658
R (all data)	0.2521	0.2926
R _{free} (all data)	0.2835	0.3267
wR ₂ (all data)	0.5012	0.5405
Largest diff. peak and hole	0.42 and −0.45 e/Å ³	0.38 and −0.37 e/Å ³
MEM Refinement		
Reflections calculated	41586	3680
Estimated Total Charge	321000	142000
R	0.050	0.054
wR	0.035	0.029

Water balance creates a threshold in soil pH at the global scale

E. W. Slessarev¹, Y. Lin², N. L. Bingham³, J. E. Johnson^{4†}, Y. Dai⁵, J. P. Schimel¹ & O. A. Chadwick³

Soil pH regulates the capacity of soils to store and supply nutrients, and thus contributes substantially to controlling productivity in terrestrial ecosystems¹. However, soil pH is not an independent regulator of soil fertility—rather, it is ultimately controlled by environmental forcing. In particular, small changes in water balance cause a steep transition from alkaline to acid soils across natural climate gradients^{2,3}. Although the processes governing this threshold in soil pH are well understood, the threshold has not been quantified at the global scale, where the influence of climate may be confounded by the effects of topography and mineralogy. Here we evaluate the global relationship between water balance and soil pH by extracting a spatially random sample ($n = 20,000$) from an extensive compilation of 60,291 soil pH measurements. We show that there is an abrupt transition from alkaline to acid soil pH that occurs at the point where mean annual precipitation begins to exceed mean annual potential evapotranspiration. We evaluate deviations from this global pattern, showing that they may result from seasonality, climate history, erosion and mineralogy. These results demonstrate that climate creates a nonlinear pattern in soil solution chemistry at the global scale; they also reveal conditions under which soils maintain pH out of equilibrium with modern climate.

Climate controls many aspects of soil chemistry, affecting soil pH (ref. 4). Alkaline soils are known to be common in arid climates, while acid soils are known to be common in humid climates¹. Surprisingly, however, the global-scale mechanisms governing this pattern remain broadly defined, and untested by direct observation. What are the dominant chemical equilibria that constrain soil pH? What aspect of climate defines the transition between alkaline and acid soils, and is the transition linear? The answers to these questions are fundamental to understanding soil development and surface geochemistry at the global scale. Furthermore, achieving this understanding may prove essential for representing soils in models of the terrestrial biosphere, given that soil pH controls many aspects of soil fertility^{5,6}. Here we illustrate that simple geochemical and hydrological concepts can be used to build a mechanistic understanding of soil pH at the global scale.

Interpretations of acid-titration experiments indicate that the soil pH is typically most strongly buffered by equilibrium with two secondary minerals: calcite (CaCO_3), or gibbsite (Al(OH)_3)^{7,8}. CaCO_3 precipitates from calcium ions (Ca^{2+}) and carbonate ions (CO_3^{2-}) derived from dissolved carbon dioxide (CO_2). Al(OH)_3 precipitates from aluminium ions (Al^{3+}) that are released from negatively charged exchange sites⁵. Both CaCO_3 and Al(OH)_3 consume protons (H^+) when they dissolve and release H^+ when they precipitate, buffering soil pH (ref. 5). Under typical laboratory conditions, soils in equilibrium with CaCO_3 and atmospheric CO_2 have a pH of 8.2 (see Methods), while the pH of soils that contain exchangeable Al^{3+} is on average 5.1 (see Methods). The presence of CaCO_3 and Al(OH)_3 is reflected in soil pH across a wide range of CaCO_3 and exchangeable Al^{3+} concentrations (Extended Data Fig. 1).

Local studies of climate gradients have shown that the relative importance of these two buffers is determined by leaching, which removes Ca^{2+} from the soil^{2,3,8,9}. In climates where evaporative demand exceeds precipitation, leaching rates are low, and dissolved Ca^{2+} accumulates as CaCO_3 —buffering soil pH near 8.2 (ref. 4). Conversely, in climates where precipitation exceeds evaporative demand, water leaches through the soil, removing Ca^{2+} and allowing accumulation of relatively immobile Al^{3+} —buffering soil pH near 5.1 (ref. 4). Because runoff and leaching rates increase abruptly as precipitation exceeds evaporative demand¹⁰, the transition between CaCO_3 and Al(OH)_3 buffered conditions is expected to occur over a small range of climatic forcing, creating a steep threshold in soil pH at the transition point between arid and humid climates^{4,8}.

However, whereas leaching controls the loss rate of Ca^{2+} , topography and mineralogy control the supply rate of Ca^{2+} to the soil solution via erosion and weathering¹¹, and thus interact with climate to influence soil pH over long timescales⁸. For instance, calcium-containing minerals may be terminally depleted in old, low-relief landscapes that have been leached in the past, limiting Ca^{2+} supply to the soil solution and creating Al(OH)_3 -buffered soils under arid conditions⁹. Alternatively, soils with short residence times in steep landscapes or areas dominated by Ca-rich rock can be rapidly supplied with Ca^{2+} from weathering, counteracting the accumulation of exchangeable Al^{3+} (ref. 12). Variation in the Ca^{2+} supply rate is challenging to constrain at global scales, and might obscure the fundamental relationship between climate and soil pH. Thus, it is unclear whether the threshold in soil pH predicted by theory prevails globally.

Nonetheless, we can search for the pH threshold at the global scale, given sufficiently extensive sampling. Statistically derived soil maps provide a tempting tool for validation^{13,14}. However, these maps rely on spatial projections of soil taxonomy that are sometimes explicitly defined by climate¹⁵, and would provide circular evidence. Thus, to test our hypotheses, we used actual measurements sampled from public databases of soil profiles (Extended Data Table 1). We then focused on pH in the subsoil (assigned here as soil to a depth of 0.5 m), to avoid effects of land-use and vegetation that might obscure the underlying geochemical signal. To overcome spatial biases in the databases (for example, heavy sampling in the USA), we developed a simple re-sampling approach that selects soil profiles randomly with respect to geographic space (Extended Data Fig. 2). We then associated these pH measurements with 1° gridded estimates of mean annual precipitation (MAP)¹⁶ and a model of potential evapotranspiration (PET)¹⁷, which represents evaporative demand. This allowed us to separate water-limited climates where leaching rates are low (MAP minus PET < 0) from energy-limited climates where leaching rates are high (MAP minus PET > 0).

Globally, the relationship between soil pH and MAP minus PET conforms closely to predictions. Soil pH at 0.5 m depth has two modes that approximate 8.2 and 5.1, the values associated with CaCO_3 and Al(OH)_3 buffers (Fig. 1). Where MAP minus PET approaches 0, there

¹Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, California, USA. ²Department of Environmental Science, Policy and Management, University of California, Berkeley, California, USA. ³Department of Geography, University of California, Santa Barbara, California, USA. ⁴Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona, USA. ⁵School of Atmospheric Sciences, Sun Yat-Sen University, Guangzhou, Guangdong, China. [†]Present address: Department of Global Ecology, Carnegie Institution for Science, Stanford, California, USA.

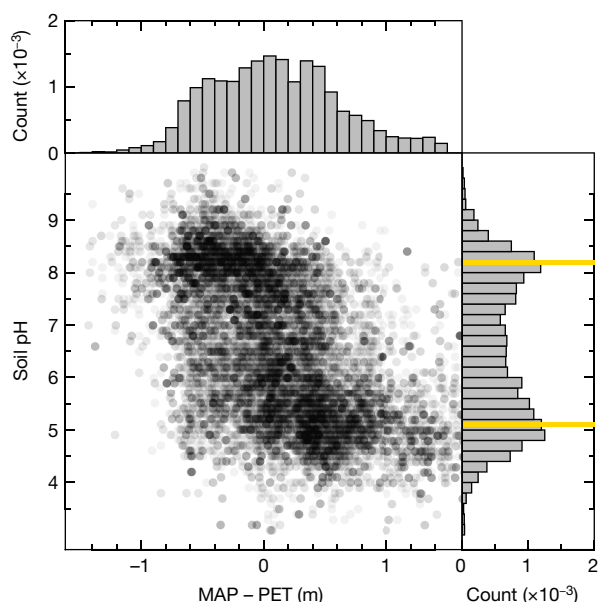


Figure 1 | Soil pH at 0.5 m depth versus annual water balance.

Transparent points show a spatial sample of 20,000 measurements of soil pH at 0.5 m depth. Side panels show histograms of MAP minus PET and soil pH, and yellow lines show the predicted pH values of CaCO_3 -buffered soils (8.2) and $\text{Al}(\text{OH})_3$ -buffered soils (5.1).

is a steep threshold between the two modes (Fig. 1). If we predict that leaching drives an immediate transition between CaCO_3 and $\text{Al}(\text{OH})_3$ equilibria where MAP minus PET = 0, the predictions explain 42% (interquartile range 42%–45%) of the observed variation in pH (here ‘variation’ means the median absolute difference from the median; see Methods). The threshold pattern is robust, appearing when MAP is balanced with an ensemble estimate of actual evapotranspiration rates¹⁸, or with simpler models of PET driven by different environmental data sets (Extended Data Fig. 3). Soil pH is not as strongly bimodal in surface soil—which may be affected by organic matter, biological cycling of Ca^{2+} , and agricultural liming—but the fundamental nonlinear pattern is still present (Extended Data Fig. 4). Furthermore, where MAP minus PET = 0, observed CaCO_3 concentrations diminish while exchangeable Al^{3+} increases, supporting the hypothesis that leaching drives a steep transition from carbonate to aluminium buffering at the global scale (Extended Data Fig. 5).

There are intriguing deviations from the central pattern—notably where acid soils occur in arid climates. These soils cluster into five regions: (1) the Sahel, (2) southern Africa, (3) northeastern Brazil, (4) Australia, and (5) mountains in western North America (Fig. 2). These acid soils may form where MAP minus PET is negative, but where appreciable leaching still occurs because of seasonal rainfall¹⁹ or snowmelt³. Additionally, geologic constraints on Ca^{2+} supply may explain acid soils in regions (1) to (4), which are low-relief continental surfaces where erosion is probably limited, and where conditions were more humid during the Last Glacial Maximum^{20–23}. Although the role of palaeoclimate in creating these acid soils is challenging to evaluate quantitatively, they are generally most common in both seasonal and low-relief environments (Extended Data Fig. 6). The prevalence of acid soils in arid, low-relief landscapes is consistent with the idea that depletion of Ca-bearing minerals might irreversibly constrain pH over long timescales, even in dry climates⁹.

Fewer deviations towards alkaline pH exist in humid climates. However, some humid soils have pH values exceeding 6.5. These measurements are scattered across several regions, including (6) southern China, (7) northern and central Europe, (8) northeastern North America, and (9) and (10) the Pacific Rim (Fig. 2). In regions (6) to (8), carbonate rocks are a major component of bedrock lithology²⁴, and might prevent soil acidification by sustaining Ca^{2+} supply to the soil solution. Comparison with a global lithologic map²⁵ shows that soil profiles in the wettest quartile of MAP minus PET are 2.6 times more likely to have a pH value > 6.5 when they fall within 1° grid cells containing carbonate bedrock (Extended Data Fig. 7). In regions (9) and (10), active volcanoes may produce easily weathered silicate minerals that could buffer pH outside the range of $\text{Al}(\text{OH})_3$ equilibrium²⁶. More generally, we observe that humid-climate soils are less acidic in high-relief landscapes (Extended Data Fig. 7), where high soil production rates may increase the availability of fresh Ca-containing minerals, increasing Ca^{2+} supply to the soil solution and thus counteracting accumulation of exchangeable Al^{3+} .

Intriguingly, the bimodal shape of the soil pH distribution indicates that soils in the neutral pH range (pH 6–7) are uncommon relative to soils in the CaCO_3 and $\text{Al}(\text{OH})_3$ buffered ranges. Soils in this pH region are thought to be buffered by mineral weathering reactions²⁷. In theory, the capacity of these reactions to neutralize H^+ is limited by the relatively slow kinetics of primary mineral dissolution, and so neutral-range soils may evolve towards CaCO_3 and $\text{Al}(\text{OH})_3$ equilibria over time⁸. Not coincidentally, neutral-range soils are intensively cultivated, because they cluster in sub-humid climates with sufficient rainfall

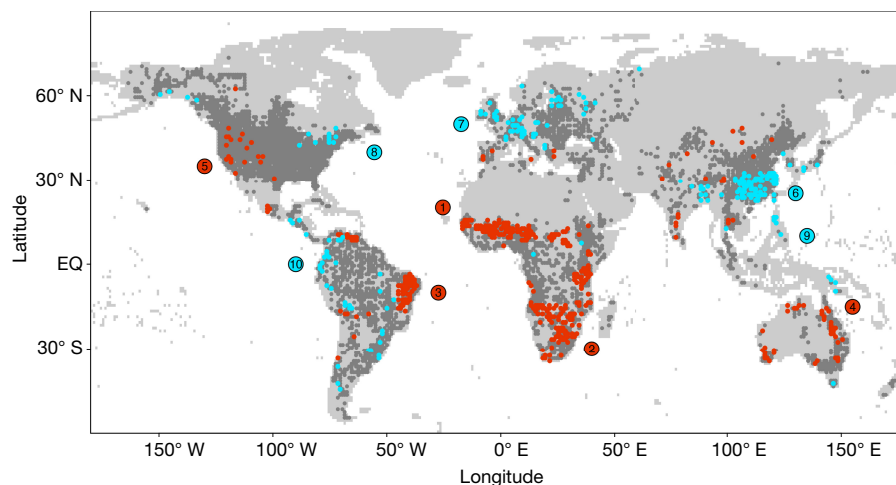


Figure 2 | Outliers from the global relationship between pH and MAP minus PET. Points show the centres of 1° grid cells containing soil profiles ($n = 4,488$). Cells in the driest quartile of MAP minus PET (in which the majority of profiles have pH < 6.5) are plotted in red, whereas cells in the

wettest quartile of MAP minus PET (in which the majority of profiles have pH > 6.5) are plotted in blue, with the remaining cells shown in grey. Numbered regions are listed in the text.

for agriculture, but retain nutrients more effectively than acid soils²⁸. Thus, from an observational standpoint, the most naturally fertile soils are relatively uncommon—and hypothetically, their relatively low prevalence may result from intrinsic aspects of their chemistry.

By assuming a threshold between CaCO₃-buffered and Al(OH)₃-buffered domains where MAP = PET, we can explain 42% (interquartile range 42%–45%) of the global variation in soil pH. The strength of this pattern indicates that a small number of specific chemical and physical mechanisms govern soil pH at the global scale. Moreover, by using this pattern as a guide, we can identify soils that appear out of equilibrium with modern climate. The distribution of these soils suggests a range of new questions that apply to the timescales of soil development: are acid soils in arid, low-relief environments irreversibly leached? Can erosion maintain high pH at a steady state in humid climates? And are neutral-range soils less common because they are poorly buffered? The answers to these questions are relevant at the timescale of human societies. Rapid changes in water balance caused by climate or land-use change might leave an increasing number of soils out of equilibrium with climate, with unknown consequences for their capacity to support productivity in natural and managed ecosystems.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 April; accepted 22 September 2016.

Published online 21 November 2016.

- Brady, N. C. & Weil, R. R. in *The Nature and Properties of Soils* 15th edn, 375–419 (Pearson Education, 2002).
- Chadwick, O. A. *et al.* The impact of climate on the biogeochemical functioning of volcanic soils. *Chem. Geol.* **202**, 195–223 (2003).
- Dahlgren, R. A., Boettinger, J. L., Huntington, G. L. & Amundson, R. G. Soil development along an elevational transect in the western Sierra Nevada, California. *Geoderma* **78**, 207–236 (1997).
- Jenny, H. in *Factors of Soil Formation: A System of Quantitative Pedology* 104–196 (Dover, 1994).
- Sposito, G. in *The Chemistry of Soils* 2nd edn, 275–295 (Oxford, 2008).
- George, E., Horst, W. & Neumann, E. in *Mineral Nutrition of Higher Plants* (ed. Marschner, P.) 409–455 (Elsevier, 2012).
- van Breemen, N., Mulder, J. & Driscoll, C. T. Acidification and alkalization of soils. *Plant Soil* **75**, 283–308 (1983).
- Chadwick, O. A. & Chorover, J. The chemistry of pedogenic thresholds. *Geoderma* **100**, 321–353 (2001).
- Vitousek, P. M. & Chadwick, O. A. Pedogenic thresholds and soil process domains in basalt-derived soils. *Ecosystems* **16**, 1379–1395 (2013).
- Budyko, M. I. in *Climate and Life* 317–370 (Academic Press, 1974).
- Hilley, G. E., Chamberlain, C. P., Moon, S., Porder, S. & Willet, S. D. Competition between erosion and reaction kinetics in controlling silicate-weathering rates. *Earth Planet. Sci. Lett.* **293**, 191–199 (2010).
- Porder, S. *et al.* Linking geomorphology, weathering and cation availability in the Luquillo Mountains of Puerto Rico. *Geoderma* **249/250**, 100–110 (2015).
- Batjes, N. H. Harmonised soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. *Geoderma* **269**, 61–68 (2016).
- Shangguan, W., Dai, Y., Duan, Q., Liu, B. & Yuan, H. A global soil data set for Earth System Modeling. *J. Adv. Model. Earth Syst.* **6**, 249–263 (2014).
- FAO-UNESCO. *FAO Soil Map of the World* <http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/faounesco-soil-map-of-the-world/en/> (The Food and Agriculture Organization of the United Nations, 1998).
- Schneider, U. *et al.* GPCC Full Data Reanalysis Version 7.0 at 1.0°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data http://doi.org/10.5676/DWD_GPCC/FD_M_V7_050 (2015).
- Leuning, R., Zhang, Y. Q., Rajaud, A., Cleugh, H. & Tu, K. A simple surface conductance model to estimate regional evaporation using MODIS leaf area index and the Penman–Monteith equation. *Wat. Resour. Res.* **44**, 1–17 (2008).
- Mueller, B. *et al.* Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis. *Hydrol. Earth Syst. Sci.* **17**, 3707–3720 (2013).
- Milly, P. Climate, interseasonal storage of soil water, and the annual water balance. *Adv. Water Resour.* **17**, 19–24 (1994).
- Kutzbach, J., Bonan, G., Foley, J. & Harrison, S. P. Vegetation and soil feedbacks on the response of the African monsoon to orbital forcing in the early to middle Holocene. *Nature* **384**, 623–626 (1996).
- Fitzsimmons, K. E. *et al.* Late Quaternary paleoenvironmental change in the Australian drylands. *Quat. Sci. Rev.* **74**, 78–96 (2013).
- Auler, A. S. & Smart, P. L. Late Quaternary paleoclimate in semiarid northeastern Brazil from U-series dating of travertine and water-table speleothems. *Quat. Res.* **55**, 159–167 (2001).
- Chase, B. M. & Meadows, M. E. Late Quaternary dynamics of southern Africa's winter rainfall zone. *Earth Sci. Rev.* **84**, 103–138 (2007).
- Ford, D. & Williams, P. in *Karst Hydrogeology and Geomorphology* 1–8 (Wiley, 2007).
- Hartmann, J. & Moosdorf, N. The new global lithological map database GLiM: a representation of rock properties at the Earth surface. *Geochem. Geophys. Geosyst.* **13**, Q12004 (2012).
- Nanzio, M., Dahlgren, R. & Shoji, S. in *Volcanic Ash Soils: Genesis, Properties and Utilization* 145–188 (Elsevier, 1993).
- Ulrich, B. Natural and anthropogenic components of soil acidification. *Z. Pflanzenernährung Bodenkunde* **149**, 702–717 (1986).
- Vitousek, P. M. *et al.* Soils, agriculture, and society in precontact Hawai'i. *Science* **304**, 1665–1669 (2004).

Acknowledgements We thank P. Vitousek, S. Fendorf, X. Feng and C. Kouba for guidance and comments. Soil data were provided by multiple contributing organizations (Extended Data Table 1). Funding for E.W.S. was provided by a Graduate Research Fellowship from the United States National Science Foundation.

Author Contributions Research was conceived by E.W.S., O.A.C., Y.L., N.L.B., J.P.S. and J.E.J. Data aggregation and processing tasks were shared by E.W.S., Y.L. and Y.D. Statistical analysis and chemical calculations were performed by E.W.S. The manuscript was written by E.W.S. with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.W.S. (eric.slessarev@lifesci.ucsb.edu).

Reviewer Information Nature thanks R. Merckx, S. Porder and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Evaporation models. We use potential evapotranspiration (PET) to represent the evaporative component of climate rather than actual evapotranspiration (AET), because PET is independent of precipitation, and thus carries more information about arid climates. Specifically, we reason that climates that are close to the MAP = PET transition are more likely to have been leached in the recent past than climates where PET greatly exceeds MAP, even if both climates have comparably small modern values of MAP minus AET. In this sense, modern PET may provide a better index of long-term leaching rates than modern AET.

We represented PET using two contrasting approaches. In the first approach, we represented PET using a modified Penman–Monteith–Leuning model, which estimates evaporation as a function of net radiation (R_n), air temperature, vapour-pressure deficit, and the aerodynamic and surface conductance of vegetation¹⁷. This approach is biophysically detailed, but it requires many parameters. Thus, in the second approach, we represented PET using the comparatively simple Priestley–Taylor equation, which models evaporative demand as a function of net radiation, air temperature, and a scaling parameter, α (ref. 29). We also explored the relationship between soil pH and the difference of MAP and AET, which we represented using the mean of the diagnostic data included in the LandFlux-EVAL synthesis¹⁸. By definition, MAP minus AET cannot take negative values, but approaches zero where PET > MAP¹⁰. We report values of MAP minus AET without imposing this constraint, and so some modelled AET values exceed MAP, resulting in slightly negative values.

Modified Penman–Monteith–Leuning model. The Penman–Monteith–Leuning model¹⁷ partitions evaporation from the plant canopy (E_c) and soil (E_s). E_c is estimated using the Penman–Monteith equation³⁰, while evaporation from soil is assumed to equal the equilibrium rate, modified by a moisture constraint. Because we were interested in obtaining an estimate of PET, we did not include a soil moisture constraint on evaporation, and then assumed that PET was equal to the sum of canopy and soil evaporation. Evaporation from wet soil can be approximated by multiplying the equilibrium evaporation rate by the Priestley–Taylor coefficient, α (ref. 31). Thus, we substituted the Priestley–Taylor model for the equilibrium model to represent soil evaporation in the Penman–Monteith–Leuning formula. The combined evaporation from canopy and soil are given by the equation:

$$\lambda E_{\text{tot}} = (sA_c + \rho c_p D_a G_a) / (s + \gamma(1 + G_a/G_c)) + \alpha(sA_s) / (s + \gamma) \quad (1)$$

where A_c and A_s are the available energy absorbed by canopy and soil (R_n minus soil heat flux, in units of $\text{MJ m}^{-2} \text{d}^{-1}$), λ is the latent heat of vaporization of water (MJ kg^{-1}), s is the slope of the saturation vapour pressure curve ($\text{kPa } ^\circ\text{C}^{-1}$), ρ is the density of air (kg m^{-3}), c_p is the specific heat of air at constant pressure ($\text{MJ kg}^{-1} ^\circ\text{C}^{-1}$), D_a is the vapour pressure deficit (kPa), γ is the psychrometric constant ($\text{kPa } ^\circ\text{C}^{-1}$), G_a is the aerodynamic conductance (m d^{-1}), and G_c is the canopy conductance (m d^{-1}). Radiation is partitioned between canopy and soil by two equations¹⁷:

$$A_c = A_{\text{tot}}(1 - e^{-k_a L}) \quad (2)$$

$$A_s = A_{\text{tot}} - A_c \quad (3)$$

where A_{tot} is equal to R_n (the soil heat flux is assumed to be negligible), L is the leaf area index, and k_a is an extinction coefficient. Canopy conductance (G_c) is constrained by maximum stomatal conductance (g_{sx}), and modified by factors that represent dependence on light availability and vapour-pressure deficit¹⁷:

$$G_c = (g_{sx}/k_Q) \ln[(Q_h + Q_{50}) / (Q_h \exp(-k_Q L) + Q_{50})] [1 / (1 + D_a/D_{50})] \quad (4)$$

where Q_h is photosynthetically active radiation at the top of the canopy (half of the incoming shortwave radiation), Q_{50} is a half-saturation constant for Q_h , D_{50} is a half-saturation constant for vapour-pressure deficit, and k_Q is the extinction coefficient for short-wave radiation.

Most parameters were obtained from a regional implementation of the Penman–Monteith–Leuning model³². The parameters k_a and k_Q were both set equal to 0.6 m^{-1} , while Q_{50} and D_{50} were set equal to $2.6 \text{ MJ m}^{-2} \text{d}^{-1}$ and 0.8 kPa (ref. 32). The maximum stomatal conductance, g_{sx} , was set equal to 0.006 m s^{-1} , which is a reasonable mean estimate for natural vegetation³³, and scaled to a daily time step. The aerodynamic conductance, G_a , is influenced by windspeed and vegetation height. Because reliable maps of both these parameters are unavailable at a global scale, we used biome-specific parameters³², assigning forests and savannas a value of 0.033 m s^{-1} , shrublands a value of 0.0125 m s^{-1} , and grasslands, cropland, and barren areas 0.01 m s^{-1} . All other parameters were calculated or obtained from the Food and Agriculture Organization (FAO) guidelines³⁴.

Priestley–Taylor model. The Priestley–Taylor model for PET uses a single parameter, α , to account for adiabatic component of latent heat transfer²⁹. While

α may vary as a function of meteorological conditions³⁰, a standard α value of 1.26 has been applied successfully at large scales³⁵. Priestley–Taylor PET is given by the equation:

$$E_{\text{tot}} = \alpha s A / (s + \gamma) \quad (5)$$

where A is total available energy (equal to R_n) and $\alpha = 1.26$. Other parameters are listed above.

Precipitation dataset. We estimated mean annual precipitation (MAP) using a 1° gridded map created from the Global Precipitation Climatology Center Full Data Reanalysis, Version 7.0¹⁶. MAP was calculated as the mean annual sum of monthly precipitation values for the years 1961–2001. We use this 40-year interval because it includes a high spatial coverage of rain-gauge stations³⁶. We corrected for systematic rain gauge measurement error using static monthly under-catch corrections³⁷ provided by the Global Precipitation Climatology Center.

Driving data for PET. Both Penman–Monteith–Leuning and Priestley–Taylor models require monthly estimates of R_n and air temperature, and the Penman–Monteith–Leuning model requires monthly estimates of vapour pressure, atmospheric pressure, surface short-wave radiation, leaf-area index, and land cover type (Extended Data Table 2). For both approaches, environmental variables obtained for multi-year time series were collapsed to monthly means of daily values before calculation of PET. PET was scaled from daily to monthly values, then summed to obtain annual PET. To test the sensitivity of our results to driving data, we used two radiation data sets: mean monthly values from the NASA/CERES energy-balanced and filled surface radiation budget, version 2.8, over the years 2001–2014³⁸, and mean monthly values from the NASA/GEWEX surface radiation budget version 3.0 over the years 1984–2007³⁹. We obtained mean monthly values of air temperature and vapour pressure from the CRU TS3.13 data set, a gridded climatology at 0.5° resolution interpolated from weather station measurements, which we averaged over the period 1961–2001, the period of maximum weather station coverage⁴⁰. Atmospheric pressure was obtained using mean elevations from the ETOPO1 global digital elevation model⁴¹ in each 1° cell and correcting using the ideal gas law³⁴. Land cover classes were obtained from the NASA MODIS satellite product MOD12⁴² and monthly means of leaf area index for the period 2001–2012 were obtained from the MODIS-derived Global Land Surface Satellite leaf area index data set^{43,44}, averaged over the period 2001–2012. All data at a higher resolution than 1° were aggregated to mean values at 1° resolution before calculation of PET.

Rainfall seasonality. We quantified rainfall seasonality by computing the coefficient of variation of under-catch corrected monthly rainfall values from the Global Precipitation Climatology Center data set.

Local relief. We estimated local topographic relief from the 1-arcminute resolution ETOPO1 digital elevation model⁴¹. Local relief was calculated as the difference between maximum and minimum elevations within a 10-km radius of each 1-arcminute cell centre. Local relief at 1° resolution was then calculated as the median relief within each 1° cell.

Carbonate lithology. We represented the extent of carbonate lithology using the Global Lithologic Map (GLiM)²⁵. We determined which 1° grid cells contained carbonate rocks by subsetting the 0.5° raster version of GLiM for carbonate lithology, and then identifying all 1° cells that contained at least one 0.5° cell classified as carbonate rock.

Soil profile data. We combined data from eight soil profile databases (Extended Data Table 1)^{45–48}. Profiles were included if they were non-duplicated and included measurements of pH in soil-water suspension. We used pH in water rather than pH in CaCl_2 or KCl solutions because pH in water is reported at a much higher frequency than pH in salt solutions. Data at 0.5 m and 0.1 m depth were obtained by selecting the horizon of each profile intersected by the corresponding depth. We selected absolute depths at 0.5 m and 0.1 m rather than soil horizons because horizon nomenclature varied across data sets. Although the choice of depths is somewhat arbitrary, the depths were selected to span the depths at which biological cycling typically influences cation concentrations⁴⁹. Using the National Cooperative Soil Survey (NCSS) database as a reference, 0.5 m approximates the median value for the top of the B horizon (0.52 m) and 0.1 m approximates the median value for the midpoint of the A horizon (0.09 m). The total number of profiles included was 60,291 at 0.5 m depth and 67,900 at 0.1 m depth (Extended Data Table 1).

Dilution ratio correction. The soil-to-water ratio of the slurry used to measure soil pH varied across data sets. To account for the effects of the soil-to-water ratio, data reported for a 1:5 ratio were corrected to a 1:1 ratio using linear correction factors⁵⁰. We could not obtain correction factors for data measured at a ratio of 1:2.5, and so left these data uncorrected. Including uncorrected data is unlikely to drive large errors in the global pH distribution because changing the soil-to-water ratio from 1:1 to 1:5 shifts pH by about 0.5 units⁵⁰, which is small relative to the

global range of soil pH values. Data measured in water without a reported ratio were assumed to be measured at ratios of 1:1 or 1:2.5.

Statistical analyses. Soil profile data were spatially resampled. In this approach, individual soil profiles were selected based on proximity to randomly distributed sampling nodes ($n = 20,000$). Sampling nodes were drawn from grid-cell centres at 1° resolution, with sampling weights based on cell area and allowing replacement. Nodes that were more than 100 km from a soil profile were not sampled to minimize edge biases. Soil profiles were selected by identifying the closest grid cell to each node that contained profiles, and then randomly drawing a profile from the total set of profiles in the cell. By design, this approach includes individual profiles multiple times in the resampled data set, with the consequence that geographically isolated profiles are included more frequently than profiles in densely sampled areas. This approach has no statistical derivation, but it produces sampling distributions that appear less-biased than the underlying data (Extended Data Fig. 1).

Water-balance model evaluation. To evaluate the relationship between MAP minus PET and soil pH, we compared observations to theoretical predictions based on calcite and gibbsite buffering systems. For all soils in grid cells where MAP minus PET < 0 , the predicted pH was 8.2, and for all remaining profiles, the predicted pH was 5.1. Residuals from the model were then computed by subtracting predicted values from observed values. Because the data are bimodally distributed, residuals from this model have a heavy-tailed distribution, and measures of variation based on squared errors (for example, the coefficient of determination, R^2) are inappropriate⁵¹. Instead, we estimated variation in the data using a robust measure of dispersion, the median absolute difference from the median (MAD). We then gauged model fit by comparing the MAD of the residuals to the MAD of the data: the percentage variation explained was equal to 1 minus MAD_{residuals}/MAD_{data}. This metric is analogous to R^2 , but makes no assumption about the distribution of the data or residuals. We estimated the uncertainty in the percentage of variation explained by resampling the data with replacement 10,000 times⁵² and calculating the interquartile range of the resulting distribution of parameter estimates.

Logistic regression of outliers. We defined 'outliers' as soils with pH < 6.5 in strongly arid climates (driest quartile of MAP minus PET) and soils with pH > 6.5 in strongly humid climates (wettest quartile of MAP minus PET). We deliberately reduced pH to this categorical expression to emphasize large-scale deviations between pH modes, rather than small-scale deviations around each mode. To quantify the prevalence of outliers as a function of rainfall seasonality, carbonate lithology, and topographic relief, we fitted logistic regressions^{53,54}. Likelihood ratio tests were used to compare regressions against the null hypothesis that the proportion of outliers is uniform with respect to each predictor⁵³. We ruled out possible collinearity between environmental predictors by checking individual correlations between predictors in both wet and dry climates. No two predictors had correlation coefficients above 0.25, and so we assume that the patterns presented are independent.

Calcite and aluminium chemistry. We used the NCSS database to validate chemical calculations and determine the relationship between climate, calcite (CaCO_3), and exchangeable aluminium (Al_x). We used the NCSS database for this purpose because it contains a large number of measurements of CaCO_3 and Al_x using consistent methods⁵⁵, and it reports the effective cation exchange capacity, which is required for modelling the pH of gibbsite buffered soils. We used a spatially resampled subset of 20,000 data points for plotting relationships with the annual water balance, following the resampling method above.

Calcite buffer. The pH of a solution exposed to calcite (CaCO_3) and open to the atmosphere can be solved using an equation derived from the chemical equilibria for CaCO_3 (ref. 56):

$$0 = H^4(2K_s/K_1K_2K_H \times p_{\text{CO}_2}) + H^3 - HK_wK_1K_H \times p_{\text{CO}_2} - K_1K_2K_H \times p_{\text{CO}_2} \quad (6)$$

where H is the hydrogen ion activity in moles, K_s is the solubility constant of CaCO_3 (in units of $\text{mol}^2 \text{L}^{-2}$), K_w is the dissociation constant for water (in $\text{mol}^2 \text{L}^{-2}$), K_1 and K_2 are the first and second dissociation constants of carbonic acid (in mol L^{-1}), K_H is Henry's constant (in $\text{mol L}^{-1} \text{atm}^{-1}$), and p_{CO_2} is the partial pressure of CO_2 (in atm). We solved this equation for H^+ at 25°C and a p_{CO_2} of 3.45×10^{-4} atm using the package rootSolve⁵⁷ in R and published parameters^{58,59}. The p_{CO_2} value of 3.45×10^{-4} atm reflects p_{CO_2} imposed by laboratory measurement conditions at standard atmospheric pressure, based on the ambient CO_2 mole fraction in 1985⁶⁰, the median measurement date of the data. Older measurements made at lower atmospheric CO_2 levels may reflect a slightly higher calcite equilibrium pH (that is, the expected pH is 8.3 before 1977). Because this difference is small and the majority of measurements were taken after this date, we report model fits for a single p_{CO_2} value.

Calcite concentrations are approximate, and reported as CaCO_3 equivalents. The NCSS database reports CaCO_3 equivalents measured using a pressure calcimeter following acid dissolution, meaning that a range of carbonate minerals are

included in the estimate⁵⁵. Also, because values are reported at a precision of 1%, some soils with $< 1\%$ CaCO_3 are probably reported with zero values, even if their pH reflects buffering by CaCO_3 .

Gibbsite buffer. The pH of a solution exposed to gibbsite ($\text{Al}(\text{OH})_3$) in a soil with exchangeable aluminium (Al_x) depends on the ratio of Al_x to other exchangeable cations (Ca_x). In nature, the solubility of $\text{Al}(\text{OH})_3$ and the exchange coefficients of clays do not follow the behaviour of purified laboratory solutions, and so the relationship between Al_x , Ca_x and pH must be estimated empirically⁶¹. To derive a typical pH for $\text{Al}(\text{OH})_3$ -buffered soils, we took the mean of all measurements from the spatial sample of the NCSS database with non-zero Al_x (pH = 5.1). Additionally, to validate the theoretical relationship between Ca_x/Al_x and pH, we fitted a model to measurements from the NCSS database taken at 0.5 m depth with non-zero Al_x and effective cation exchange capacity. The Gapon exchange model can be used to develop a log-linear relationship between Ca_x/Al_x and pH (ref. 61):

$$\text{pH} = b_0 + b_1 \log_{10}(\text{Ca}_x/\text{Al}_x) \quad (7)$$

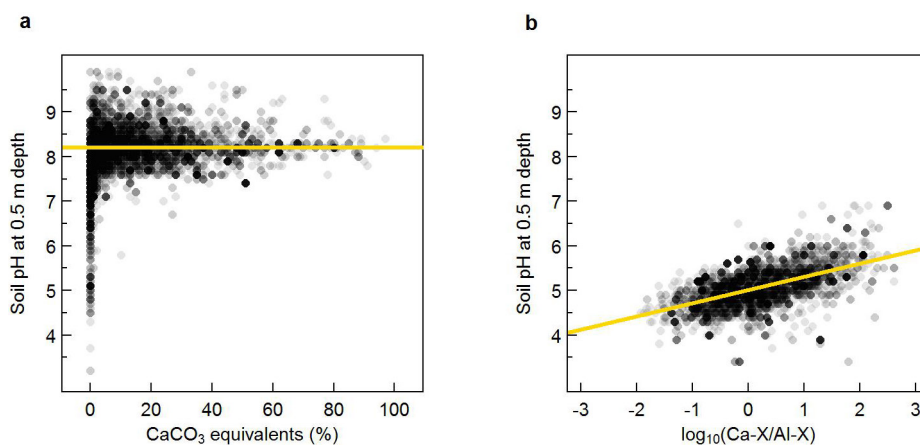
where b_0 and b_1 are fitted constants. To fit the model, we assumed that Al_x was equal to aluminium extractable in 1 M KCl, and Ca_x was equal to the effective cation exchange capacity minus Al_x . The data show a log-linear relationship between Ca_x/Al_x and pH (Extended Data Fig. 1, $b_0 = 4.96$, $b_1 = 0.32$, $R^2 = 0.36$, $P < 0.01$), supporting control of pH by Ca_x/Al_x . However, we note that the relationship appears slightly concave-curvilinear, suggesting that the Gapon model fails to account for the total activity of Al_x . This issue warrants further investigation.

Code availability. Code used to spatially resample soil profiles, calculate PET, and perform statistical analyses are maintained on GitHub and publicly exchange online at <http://dx.doi.org/10.5281/zenodo.61996>. Code for pre-processing of raw data sets is available from the authors upon request.

Data availability. All soil profile and meteorological data used in this study are publicly available from the sources listed in the text and in Extended Data Tables 1 and 2. Several of the soil profile databases are only available by direct request from the providing institutions (see Extended Data Table 1). As such, the combined soil profile data set used in this study is available from the authors upon request, given permission from providing institutions.

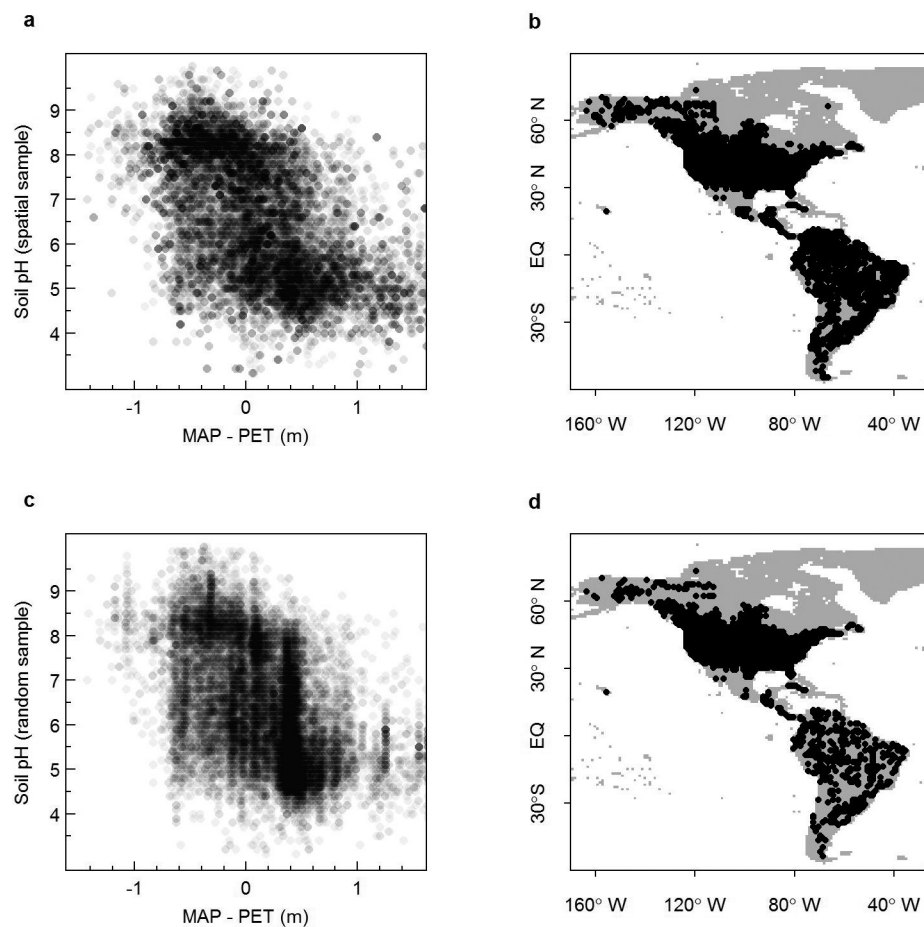
29. Priestley, C. H. B. & Taylor, R. J. On the assessment of surface heat flux and evaporation using large scale parameters. *Mon. Weath. Rev.* **100**, 81–92 (1972).
30. Monteith, J. & Unsworth, M. in *Principles of Environmental Physics* 4th edn, 217–247 (Elsevier, 2013).
31. Eichinger, W., Parlange, M. B. & Stricker, H. On the concept of equilibrium evaporation and the value of the Priestley-Taylor coefficient. *Wat. Resour. Res.* **32**, 161–164 (1996).
32. Zhang, Y. Q., Chiew, H. S., Zhang, L., Leuning, R. & Cleugh, H. A. Estimating catchment evaporation and runoff using MODIS leaf area index and the Penman Monteith equation. *Wat. Resour. Res.* **44**, W10420 (2008).
33. Kelliher, F. M., Leuning, R., Raupach, M. R. & Schulze, E. D. Maximum conductances for evaporation from global vegetation types. *Agric. For. Meteorol.* **73**, 1–16 (1995).
34. Allen, R. G., Pereira, L. S., Raes, D. & Smith, M. *Crop Evapotranspiration—Guidelines for Computing Crop Water Requirements*. FAO Irrigation and Drainage Paper 56, <http://www.fao.org/docrep/x0490e/x0490e00.htm> (FAO, 1998).
35. Fisher, J. B., Tu, K. P. & Baldocchi, D. D. Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. *Remote Sens. Environ.* **112**, 901–919 (2008).
36. Schneider, U., Becker, A. & Finger, P. GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theor. Appl. Climatol.* **115**, 15–40 (2014).
37. Legates, D. R. & Willmott, C. J. Mean seasonal and spatial variability in gauge-corrected, global precipitation. *Int. J. Climatol.* **10**, 111–127 (1990).
38. Wielicki, B. A. et al. Clouds and the Earth's Radiant Energy System (CERES): an Earth observing system experiment. *Bull. Am. Meteorol. Soc.* **77**, 853–868 (1996).
39. Stackhouse, P. W. et al. The NASA/GEWEX surface radiation budget release 3.0: 24.5-Year Dataset. *GEWEX News* **21**, 10–12 (2011).
40. Harris, I., Jones, P. D., Osborne, T. J. & Lister, D. H. Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 dataset. *Int. J. Climatol.* **34**, 623–642 (2014).
41. Amante, C. & Eakins, B. W. *ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis*. NOAA Technical Memorandum NESDIS NGDC-24, <http://www.doi.org/10.7289/V5C8276M> (National Geophysical Data Center, NOAA, 2009).
42. Friedl, M. A. et al. Global land cover mapping from MODIS: algorithms and early results. *Remote Sens. Environ.* **83**, 287–302 (2002).
43. Liang, S. & Xiao, Z. *Global Land Surface Products: Leaf Area Index Product Data Collection (1985–2010)*. <http://www.doi.org/10.6050/glass863.3004.db> (Beijing Normal University, 2012).
44. Xiao, Z. et al. Use of general regression neural networks for generating the GLASS leaf area index product from time series MODIS surface reflectance. *IEEE Trans. Geosci. Remote Sens.* **52**, 209–223 (2013).

45. Batjes, N. H. Harmonized soil profile data for applications at global and continental scales: updates to the WISE database. *Soil Use Manage.* **25**, 124–127 (2009).
46. Leenaars, J. G. B., van Oostrum, A. M. J. & Ruiperez Gonzalez, M. *Africa Soil Profiles Database, Version 1.2. A Compilation of Georeferenced and Standardized Legacy Soil Profile Data for Sub-Saharan Africa (with Dataset)*. ISRIC Report 2014/1 (Africa Soil Information Service (AfSIS) project, ISRIC, World Soil Information, 2012).
47. Hiederer, R., Jones, R. J. & Daroussin, J. Soil Profile Analytical Database for Europe (SPADE): reconstruction and validation of the measured data (SPADE/M). *Geograf. Tids. Dan. J. Geogr.* **106**, 71–85 (2006).
48. Cooper, M., Mendes, L. M. S., Silva, W. L. C. & Sparovek, G. A national soil profile database for Brazil available to international scientists. *Soil Sci. Soc. Am. J.* **69**, 649–652 (2005).
49. Jobbágy, E. G. & Jackson, R. B. The uplift of soil nutrients by plants: biogeochemical consequences across scales. *Ecology* **85**, 2380–2389 (2004).
50. Libohova, Z. *et al.* Converting pH 1:1 H₂O and 1:2 CaCl₂ to 1:5 H₂O to contribute to a harmonized global soil database. *Geoderma* **213**, 544–550 (2014).
51. Willmott, C. J., Matsuura, K. & Robeson, S. M. Ambiguities inherent in sums-of-squares-based error statistics. *Atmos. Environ.* **43**, 749–752 (2009).
52. Efron, B. Bootstrap methods: another look at the jackknife. *Ann. Stat.* **7**, 1–26 (1979).
53. Trexler, J. C. & Travis, J. Nontraditional regression analyses. *Ecology* **74**, 1629–1637 (1993).
54. R Core Team. *R: a Language and Environment for Statistical Computing*. <https://www.R-project.org/> (Foundation for Statistical Computing, 2016).
55. Soil Survey Staff. In *Kellogg Soil Survey Laboratory Methods Manual Version 5.0 Soil Survey Investigations*. Report No 42 (eds Burt, R. & Soil Survey Staff) http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/home/?cid=nrcs142p2_ (United States Department of Agriculture Natural Resources Conservation Service, 2014).
56. Plummer, N. L. & Busenberg, E. The solubilities of calcite, aragonite and vaterite in CO₂ H₂O solutions between 0 and 90°C, and an evaluation of the aqueous model for the system CaCO₃–CO₂–H₂O. *Geochim. Cosmochim. Acta* **46**, 1011–1040 (1982).
57. Soetaert, K. rootSolve: nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations. *R package version 1.6* <https://cran.r-project.org/web/packages/rootSolve/index.html> (2009).
58. Parkhurst, D. L. & Appelo, C. A. J. *Users Guide to PHREEC (Version 3.0): a Computer Program for Speciation, Batch-Reaction, One-Dimensional Transport, and Inverse Geochemical Calculations*. http://wwwbrr.cr.usgs.gov/projects/GWC_coupled/phreeqc/index.html (United States Geological Survey, 2013).
59. Sander, R. Compilation of Henry's law constants (version 4.0) for water as solvent. *Atmos. Chem. Phys.* **15**, 4399–4981 (2015).
60. Keeling, C. D. & Whorf, T. P. Atmospheric CO₂ concentrations derived from flask air samples at sites in the SIO network. In *Trends: A Compendium of Data on Global Change* http://cdiac.ornl.gov/trends/co2/sio-keel-flask/sio-keel-flaskmlo_c.html# (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, United States Department of Energy, 2004).
61. Reuss, J. O., Walthall, P. M., Roswall, E. C. & Hopper, R. W. E. Aluminum solubility, calcium-aluminum exchange, and pH in acid forest soils. *Soil Sci. Soc. Am. J.* **54**, 374–380 (1990).

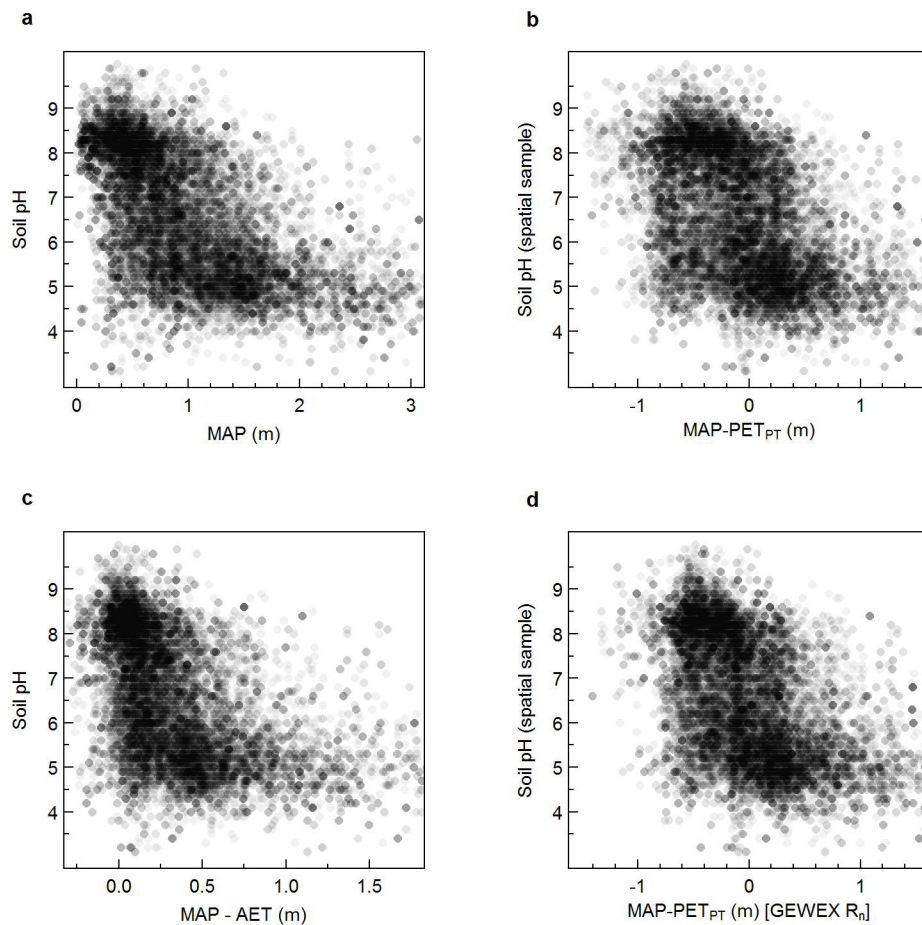


Extended Data Figure 1 | Soil pH versus calcite and exchangeable aluminium. Transparent points show a spatial sample of 20,000 measurements from the NCSS database. **a**, The relationship between soil pH at 0.5 m and CaCO_3 equivalents as a mass percentage. The yellow line shows the calculated pH of a solution in equilibrium with calcite and

atmospheric CO_2 (345 parts per million) at 25 °C. **b**, The relationship between soil pH at 0.5 m and the log-ratio of exchangeable calcium (Ca_x) to exchangeable aluminium (Al_x), which is thought to control the pH of gibbsite-buffered soils. The yellow line is the fit by least-squares regression ($b_0 = 4.96$, $b_1 = 0.32$, $R^2 = 0.36$, $P < 0.01$).

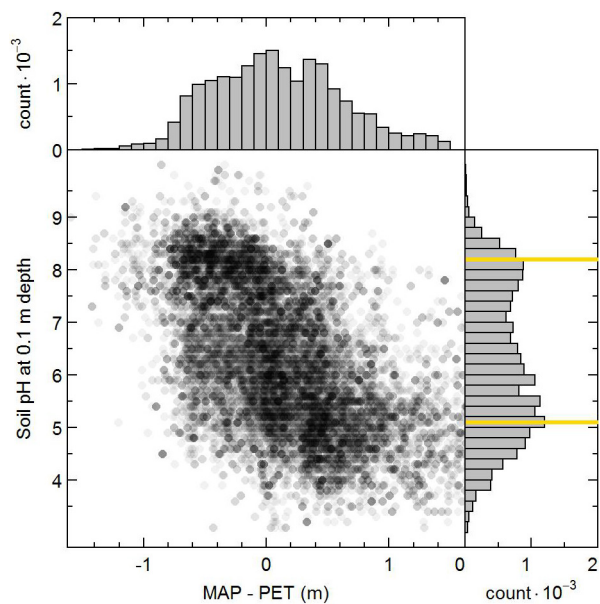


Extended Data Figure 2 | Results of spatial resampling. Transparent points show a spatial sample of 20,000 measurements (a and b) and a random sample of 20,000 measurements (c and d). **a, c,** pH at 0.5 m depth versus MAP minus PET. **b, d,** The geographic distribution of measurements in the Americas.

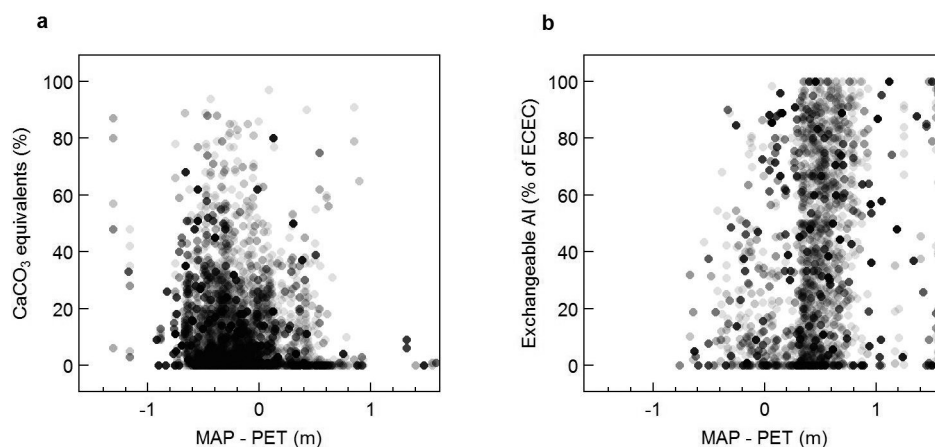


Extended Data Figure 3 | Soil pH at 0.5 m depth versus alternative water-balance models. Transparent points show a spatial sample of 20,000 measurements of soil pH at 0.5 m depth. **a**, Soil pH versus MAP. **b**, Soil pH versus MAP minus PET estimated using the Priestley–Taylor method

driven by CERES radiation data. **c**, MAP minus AET, from the LandFlux-EVAL synthesis. **d**, MAP minus PET estimated using the Priestley–Taylor method driven by GEWEX radiation data.

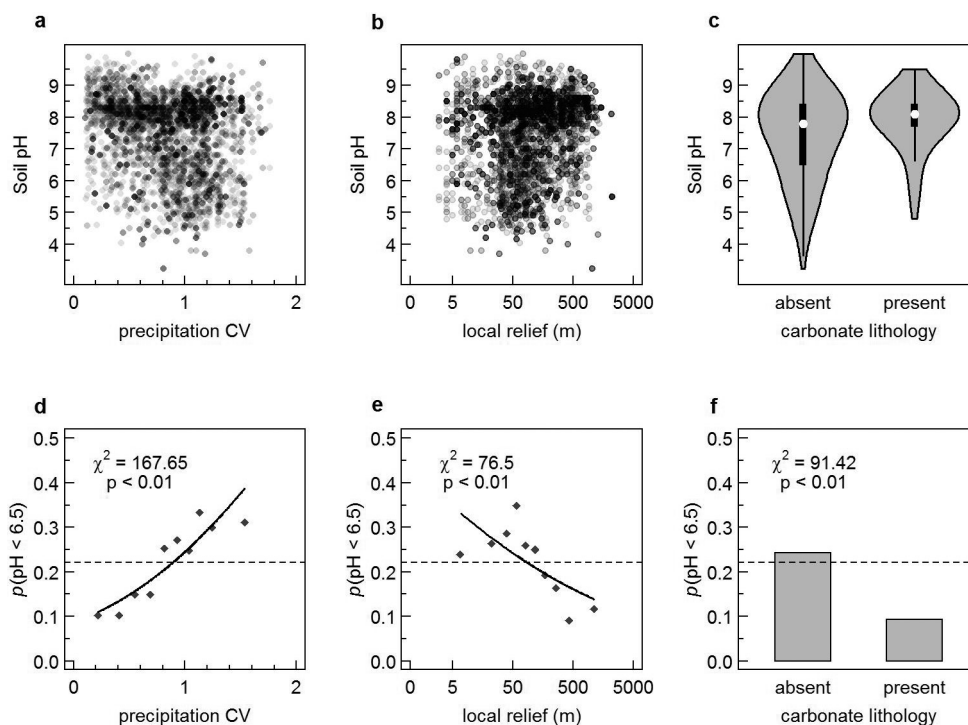


Extended Data Figure 4 | Soil pH at 0.1 m depth versus MAP minus PET. Transparent points show a spatial sample of 20,000 measurements of soil pH at 0.1 m depth. Side panels show histograms of MAP minus PET and soil pH, and yellow lines show predicted pH values of CaCO_3 -buffered soils (8.2) and $\text{Al}(\text{OH})_3$ -buffered soils (5.1).



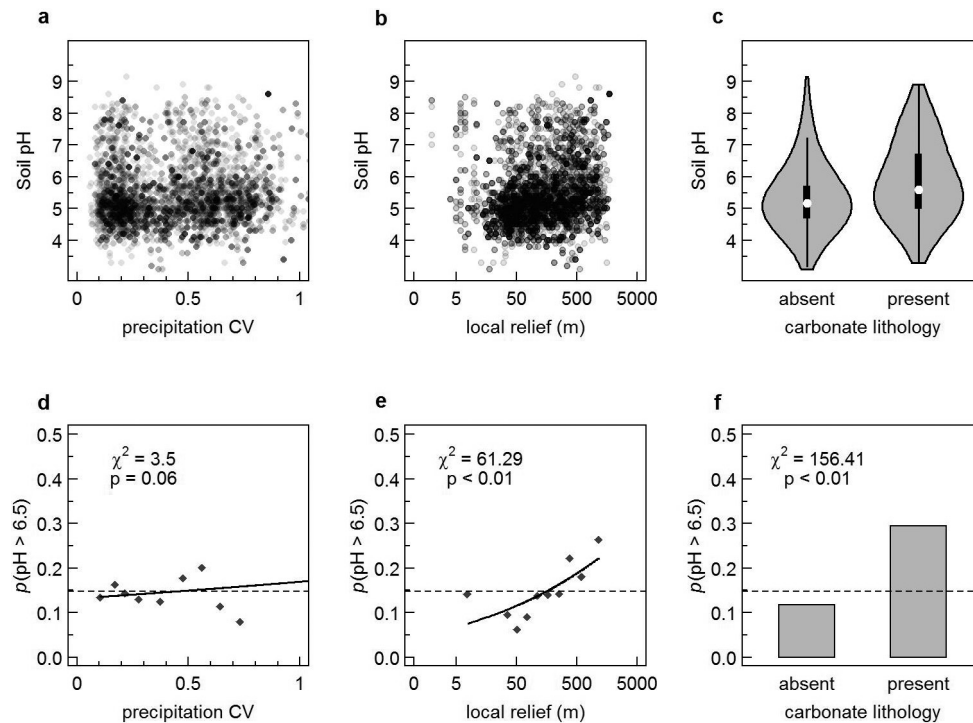
Extended Data Figure 5 | Calcite and exchangeable aluminium versus MAP minus PET. Transparent points represent a spatial sample of 20,000 measurements from the NCSS database. **a**, Calcite (CaCO₃) equivalents as mass percentage versus MAP minus PET. **b**, Exchangeable aluminium as

a percentage of the effective cation exchange capacity versus MAP minus PET. These data are not reported for all samples in the NCSS database, and so points on the plot represent only the subset of the data with reported values.



Extended Data Figure 6 | Dry-climate soil pH versus seasonality, relief and carbonates. Transparent points show soil pH at 0.5 m depth in the driest quartile of MAP minus PET ($n = 5,000$). **a**, Soil pH versus the coefficient of variation (CV) of precipitation. **b**, Soil pH versus local relief. **c**, Violin plots showing soil pH versus carbonate lithology. Panels **d** and **e** show the proportion of the observations with pH < 6.5, binned into

deciles of the variable on the x axis; panel **f** shows the proportion in each lithologic category. Black lines show logistic regression fits, with associated chi-squared (χ^2) statistics and P values from likelihood ratio tests for precipitation CV ($\chi^2 = 167.65$, $P < 0.01$), local relief ($\chi^2 = 76.5$, $P < 0.01$) and carbonate lithology ($\chi^2 = 91.42$, $P < 0.01$). Dashed lines show the proportion of observations with pH < 6.5.



Extended Data Figure 7 | Wet-climate soil pH versus seasonality, relief and carbonates. Transparent points show soil pH at 0.5 m depth in the wettest quartile of MAP minus PET ($n = 5,000$). **a**, Soil pH versus the coefficient of variation of precipitation. **b**, Soil pH versus local relief. **c**, Violin plots showing soil pH versus carbonate lithology. Panels **d** and **e** show the proportion of the observations with pH > 6.5, binned into

deciles of the variable on the x axis; panel **f** shows the proportion in each lithologic category. Black lines show logistic regression fits, with associated χ^2 statistics and P values from likelihood ratio tests for precipitation CV ($\chi^2 = 3.5$, $P = 0.06$), local relief ($\chi^2 = 61.29$, $P < 0.01$) and carbonate lithology ($\chi^2 = 156.41$, $P < 0.01$). Dashed lines show the proportion of observations with pH > 6.5.

Extended Data Table 1 | Soil profile data sets

Dataset	Provider	Reference	# profiles used in analysis
National Cooperative Soil Survey	United States Department of Agriculture Natural Resources Conservation Service	*	34,259 (0.1 m) 35,775 (0.5 m)
Chinese National Soil Database	Chinese Soil Survey	*	2,183 (0.1 m) 2,370 (0.5 m)
World Inventory of Soil Emissions- Potentials	International Soil Reference and Information Center	45	2,505 (0.1 m) 2,682 (0.5 m)
Africa Soil Profile Database	International Soil Reference and Information Center	46	10,057 (0.1 m) 11,680 (0.5 m)
Australian National Soil Database	Commonwealth Scientific and Industrial Research Organization	*	4,389 (0.1 m) 5,959 (0.5 m)
Canadian National Soil Database	Canadian Soil Information Service	*	6,035 (0.1 m) 8,219 (0.5 m)
Soil Profile Analytical Database of Europe/ Measured Parameters	European Soil Data Center	47	375 (0.1 m) 408 (0.5 m)
Brazilian National Soil Database	Luiz de Queiroz College of Agriculture	48	488 (0.1 m) 807 (0.5 m)

Data sets marked with an asterisk are publicly available on request from the data provider. The other datasets are described in refs 45–48.

Extended Data Table 2 | Gridded Environmental Data sets

Dataset	Provider	Reference
GPCC Full Data Reanalysis v7.0	Global Precipitation Climatology Center	16
LandFlux-EVAL Synthesis	Institute for Atmospheric and Climate Science, ETH Zurich	18
GLiM Global Lithologic Map	Institute for Geology, Universität Hamburg	25
CERES Surface Radiation Budget v2.8	United States National Aeronautics and Space Administration	38
GEWEX Radiation Budget v3.0	United States National Aeronautics and Space Administration	39
CRU TS3.13 Global Surface Climatology	University of East Anglia Climate Research Unit	40
ETOPO1 Digital Elevation Model	United States National Centers for Environmental Information	41
MOD12 Land Cover Classes	United States National Aeronautics and Space Administration	42
GLASS Leaf Area Index	Global Land Cover Facility	43, 44

The models are described in refs 16, 18, 25, 38–44.

Broadening not strengthening of the Agulhas Current since the early 1990s

Lisa M. Beal¹ & Shane Elipot¹

Western boundary currents—such as the Agulhas Current in the Indian Ocean—carry heat poleward, moderating Earth's climate and fuelling the mid-latitude storm tracks^{1,2}. They could exacerbate or mitigate warming and extreme weather events in the future, depending on their response to anthropogenic climate change. Climate models show an ongoing poleward expansion and intensification of the global wind systems, most robustly in the Southern Hemisphere^{3–5}, and linear dynamical theory^{6,7} suggests that western boundary currents will intensify and shift poleward as a result^{3,8}. Observational evidence of such changes comes from accelerated warming and air–sea heat flux rates within all western boundary currents, which are two or three times faster than global mean rates^{5,9,10}. Here we show that, despite these expectations, the Agulhas Current has not intensified since the early 1990s. Instead, we find that it has broadened as a result of more eddy activity. Recent analyses of other western boundary currents—the Kuroshio and East Australia currents—hint at similar trends^{11–15}. These results indicate that intensifying winds may be increasing the eddy kinetic energy of boundary currents, rather than their mean flow. This could act to decrease poleward heat transport and increase cross-frontal exchange of nutrients and pollutants between the coastal ocean and the deep ocean. Sustained *in situ* measurements are needed to properly understand the role of these current systems in a changing climate.

To estimate the trend in Agulhas Current transport we build a 22-year proxy using three years of *in situ* measurements from the Agulhas Current Time-series (ACT) array¹⁶ combined with coincident along-track satellite altimeter data spanning the years 1993–2015 (Fig. 1).

We define two measures of transport for the Agulhas Current: a streamwise, southwestward jet transport T_{jet} , and a geographically fixed, net boundary-layer transport T_{box} . Over the three years of *in situ* data the mean and standard deviation of T_{jet} are -84 Sv ($1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$) and 24 Sv , respectively, and of T_{box} are -77 Sv and 32 Sv , respectively¹⁶. In past studies T_{box} has been more often applied to quantify boundary current flow, yet this measure suffers spurious effects from meander events¹⁶, which are largely removed in the streamwise case.

Before building a proxy we first test for the necessary condition of a linear and fixed relationship between our *in situ* transports and sea surface slope from satellite altimeter. Sea surface slope is equivalent to surface geostrophic velocity, and on the basis of previous analyses we expect the relationship between surface geostrophic velocity and full-depth transport to be strong in the Agulhas Current, despite the presence of an undercurrent^{17,18}. Using empirical orthogonal function (EOF) analysis, we find that transport and sea surface height and slope along the ACT array do exhibit similar and significantly correlated modes of variance (correlations >0.7 , P values $<10^{-3}$). In each case, these modes express weakening or strengthening, broadening or narrowing, and meandering of the Agulhas Current jet (Extended Data Fig. 1 and Methods). Hence, variance of the sea surface is strongly tied to oceanic transport along the ACT array at 34° S and

using altimetry to build a proxy for Agulhas Current transport seems physically justifiable.

For our Agulhas Current transport proxy we build nine linear regression models between sea surface slope and transport per unit distance T_x , one at each of the ACT mooring locations (Fig. 1). We then fit a smoothed function at 1-km intervals and integrate horizontally to obtain T_{jet} and T_{box} for each altimeter pass. In this way, we can account for a current that meanders and changes in width by allowing the core and flanks of the flow to vary at different rates, although the vertical structure necessarily remains fixed. A proxy based on regression between array-wide sea level variance and total transport is more statistically successful for T_{box} ; however, we find that this proxy's trend is spurious because of the observed broadening of the current (Extended Data Fig. 2 and Methods).

Our 22-year proxy time series are shown in Fig. 2, together with their frequency spectra and seasonal cycles. For T_{box} our proxy explains 61% of the variance during the ACT period, while for T_{jet} it explains 55% of the variance. The jet and boundary layer transports have spectral peaks that are significant (see Methods subsection 'Spectral Estimates') at the annual period (Fig. 2c) and their seasonal cycles match those of the *in situ* data¹⁶, with weakest transports occurring during austral winter, in August, and strongest transports occurring throughout austral summer (Fig. 2d). At most timescales, the variance of T_{jet} is less than that of T_{box} (Fig. 2c and d).

The long-term trends are $+1.0 \pm 2.4 \text{ Sv per decade}$ for T_{box} and $+2.1 \pm 2.1 \text{ Sv per decade}$ for T_{jet} (where tolerances are the 95% confidence interval). These trends would correspond to a slight decrease in Agulhas Current transport, although neither trend is significant. In the Indian Ocean, a spin-up of the ocean circulation over the northern reaches of the subtropical gyre has been implied by a small intensification of the trade winds⁴ and an increase in the Indonesian through-flow¹⁹ since the early 1990s. Farther south at the ACT array the trend in wind stress curl is less certain, since changes in the strength and latitudinal position of the Westerlies differ between wind products²⁰. Nevertheless, the derived winds from the 20CRv2 reanalysis show an upward trend in wind curl over all the southern subtropical oceans²¹, and rapidly increasing sea surface temperatures and air–sea fluxes have been attributed to a strengthening and poleward shift of the Agulhas Current^{5,9,10}.

Given that our results appear to contrast with these indicators, we examine our proxy trends more closely. We look at the structure of sea level and slope trends along the ACT array (Fig. 3a and b) and compare them to our transport trend. We find a non-uniform pattern of sea level rise across the Agulhas Current, with a minimum at 130 km from the coast (Fig. 1). This pattern gives rise to a negative trend in sea surface slope across the inshore half of the boundary layer and a positive trend offshore (Fig. 3b). Since the mean sea surface slope is positive across the Agulhas Current (Extended Data Fig. 1), these trends result in weaker velocities within the core of the current and stronger velocities over its flank. The effect on the Agulhas Current is a weakening and broadening of the jet, as illustrated by the change in T_x across the array

¹Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Miami, Florida, USA.

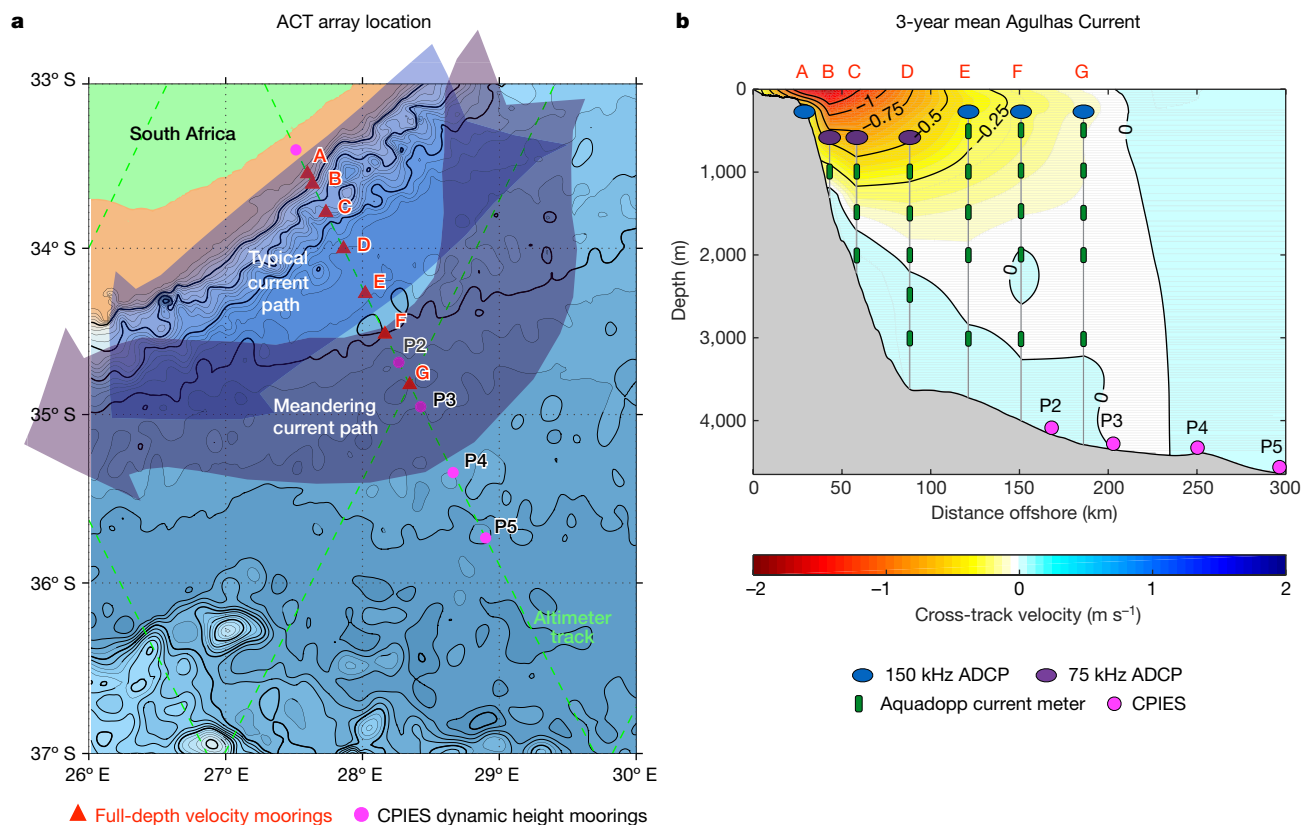


Figure 1 | Agulhas Current Time-series (ACT) instrumental array and mean Agulhas Current from April 2010 to February 2013.

a, Geographical location of ACT mooring array, consisting of current meter moorings A–G and CPIES (Current Pressure Inverted Echo Sounders) sites P2–P5. Bathymetry down to 1,000 m is shown in tan and in deepening shades of blue thereafter, with contours every 200 m and thick contours every 1,000 m, from 200 m to 5,000 m. **b**, Vertical section

of the mooring array superposed on the 3-year mean cross-track velocity. Bathymetry is shaded grey. Southwestward velocities (Agulhas Current) are shaded yellow through to red and northeastward velocities are shaded blue (see colour scale). Acoustic Doppler Current Profilers (ADCPs) measure velocity throughout the upper water column. Aquadopp current meters measure velocity at a single point. Pairs of CPIES estimate profiles of geostrophic velocity.

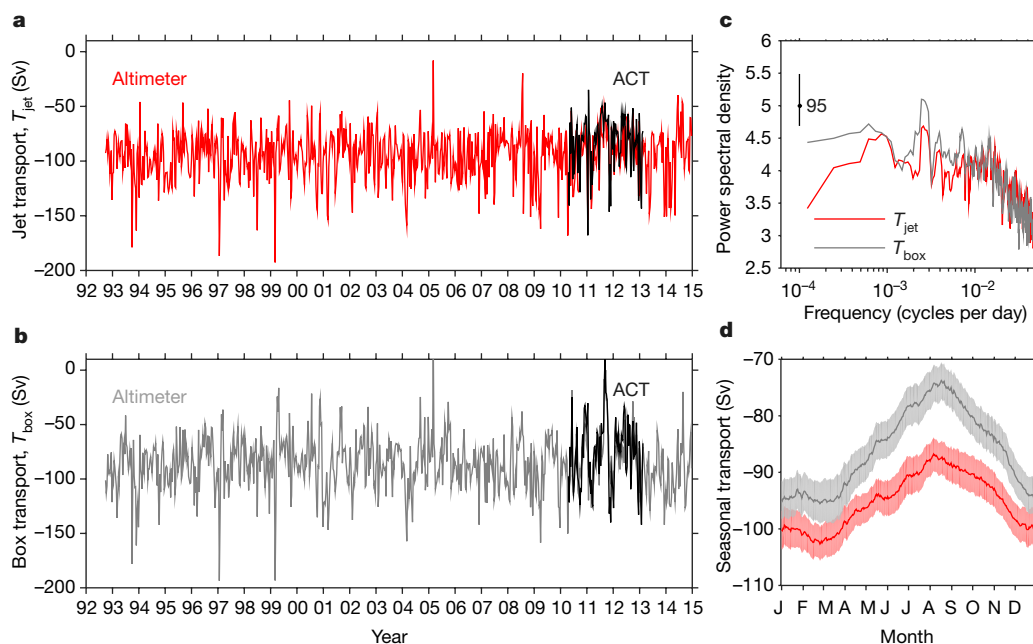


Figure 2 | Agulhas Current transport proxies based on regressions between transport and sea surface slope at each ACT mooring location.

a, Proxy time series for jet or stream-wise transport T_{jet} . **b**, Proxy time series for boundary layer transport T_{box} . The three years of *in situ* transports from the ACT array are shown as black lines. **c**, Frequency

spectra of the jet (red) and boundary layer (grey) proxies. The 95% confidence interval is shown. **d**, Seasonal transport (daily average values) of the jet (red) and boundary layer (grey) proxies. Shading shows 95% confidence intervals.

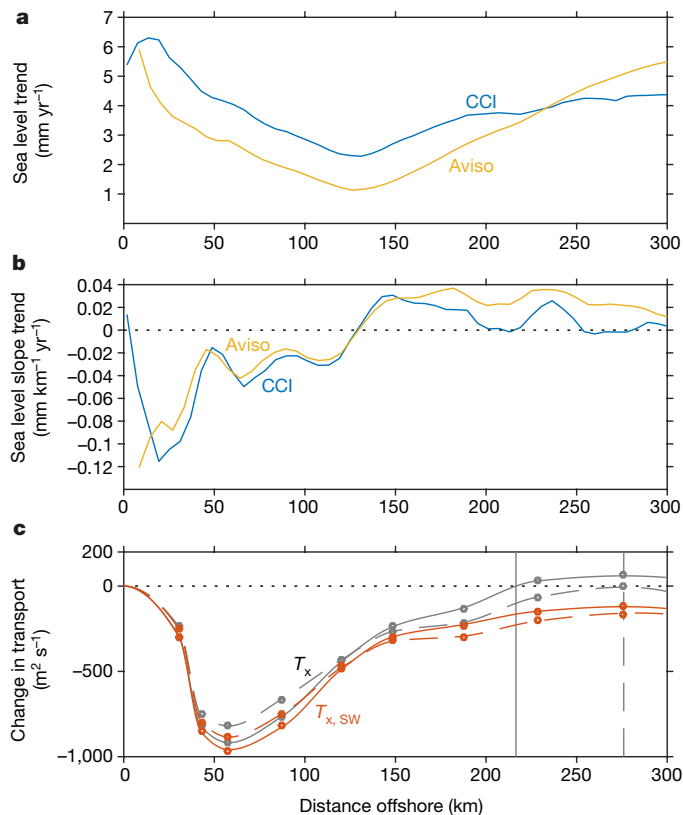


Figure 3 | Trends of sea level and oceanic transport across the Agulhas Current, showing broadening. **a**, **b**, Sea level (a) and sea surface slope (b) linear trends estimated using along-track satellite altimeter from Aviso (blue) and CCI (orange) products (1993–2015). **c**, Transport per unit distance based on regressions between sea surface slope and oceanic transport at each ACT mooring site (circles). Transport changes between the beginning (solid lines) and end (dashed lines) of the altimeter record are implied by the linear trends. T_x is net transport per unit distance and $T_{x, SW}$ is its southwestward component (as used to calculate T_{jet}). Vertical lines indicate the width of the boundary layer at the beginning (solid) and end (dashed) of the altimeter record.

from the regression models at each mooring (Fig. 3c). The core of the current is in the same position, but weaker, while flow throughout the offshore flank of the current is stronger. The resultant broadening of the boundary layer, defined by the zero crossing of T_x , is about 50 km (Fig. 3c). Hence, we conclude that the Agulhas Current is weakening and broadening over time, while its total transport remains stable.

Broadening of the Agulhas Current can be understood in the context of a simple Munk model²², whereby the width of the western boundary layer will increase if the lateral viscosity increases. This could occur as a result of an increase in eddy activity in the current over time (Fig. 4). We derive the trend in eddy kinetic energy (EKE) of the Agulhas Current at this latitude from a mapped altimeter product using a fixed number of satellites over time, as has been done previously²⁰. Consistent with this model, the trend in EKE is everywhere positive across the current, while the peak in mean kinetic energy (MKE) within the core of the jet is decreasing (Fig. 4).

Broadening can also be understood in terms of meandering. Mesoscale meanders dominate the variance—and therefore the EKE—of the Agulhas Current²³, growing largely through a barotropic conversion of energy from the mean horizontal flow^{17,18}. Such instabilities act to transfer energy offshore and decelerate the core of the jet²⁴. Hence, a reported increase in the number of meander events over time¹⁷ will lead to a weakening and broadening of the flow.

Measurements in other western boundary currents also point to trends of increasing EKE, rather than increasing transports, as we observe in the Agulhas Current. In the Pacific Ocean, there is an

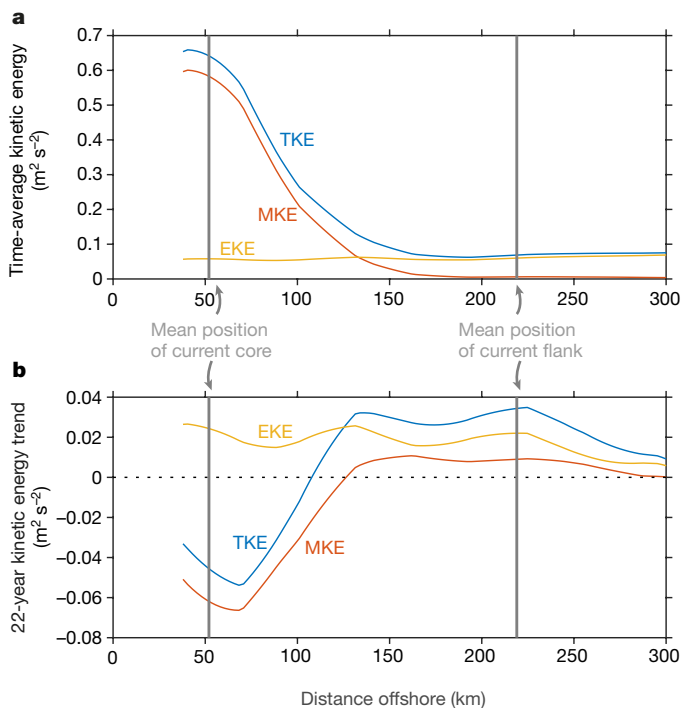


Figure 4 | Kinetic energy analyses across the Agulhas Current, showing increased eddying. **a**, Total kinetic energy (TKE), mean kinetic energy (MKE) and eddy kinetic energy (EKE) from the mapped Aviso altimeter product. Mean and eddy kinetic energies are defined as variability at timescales of less and more than eighteen months, respectively. **b**, TKE, MKE and EKE linear changes over 22 years. In **a** and **b** the mean position of the current core and its offshore flank during the three years of ACT are shown as vertical grey lines. Data within 40 km of the coast are not available in the mapped altimeter product.

increase in observed trade winds over the last twenty years, some of which is attributed to climate change²⁵. Increases in sea surface height over the north and south subtropical Pacific point to a concomitant intensification of the ocean gyres⁴. However, no compensating trends have been found in the transports of the Kuroshio or East Australia currents^{13,14}, even while they have been warming¹⁰. Instead, there is evidence that the Kuroshio Current is broadening, with the same pattern of sea level rise as seen across the Agulhas Current¹¹, and that eddy activity in both the Kuroshio and East Australia currents is increasing^{12–14}. Furthermore, eddy variability of the East Australia Current has recently been linked to regional wind stress curl¹⁵. In the Atlantic, prediction of trends in Gulf Stream intensity is complicated by a potential weakening of the overturning circulation²⁶ and by uncertain trends in wind curl¹⁰. Sea level changes along the east coast of the USA have been used to suggest a weakening of the Gulf Stream²⁷, but this is not corroborated by *in situ* measurements, which show no trend²⁸.

Extending (that is, using data to extend *in situ* observations back in time) and inferring ocean circulation changes using satellite altimeter data is becoming commonplace^{14,19,27}, in an attempt to understand oceanic change better despite a paucity of measurements. Our results call for caution when inferring trends in currents using sea surface height difference alone, since sea level changes may be inhomogeneous across the current (Fig. 3). The exact position of the current, and any broadening or narrowing over time, must be taken into account (see Methods). On multidecadal timescales, the implicit assumption of a fixed vertical stratification may also become problematic, as thermohaline changes become important. More hydrographic data are necessary, particularly within western boundary currents, to be able to estimate trends in stratification.

A particular weakness of our analysis is that it captures only changes in the Agulhas Current at 34° S, the location of the ACT array. Farther north, opposing trends in MKE and EKE are such that the mean flow

appears to be strengthening over time and there are fewer eddies²⁰. However, these inferences are made where the current core is within 40 km of the coast, a region of poor altimeter coverage, and are uncorroborated by *in situ* measurements. Another weakness is that wind products exhibit large discrepancies in the mean²⁹, such that temporal changes are even less reliable²⁰, and a poleward shift adds further uncertainty to any observed trend in wind stress curl⁵. Finally, a linear trend model accounts for only a small fraction of the variance in kinetic energy and transport of the Agulhas Current. 22 years of satellite observations are barely sufficient to discern anthropogenic trends, although decadal climate variability in the Indian Ocean is smaller than in the Pacific and Atlantic oceans. 50-year trends in sea surface temperature in the Indian Ocean sector are consistent with trends over the last two decades⁵.

Our results, together with recent analyses in other western boundary currents, suggest that intensifying winds may act to increase the EKE of boundary currents, rather than their mean flow. This hypothesis draws parallels with the eddy compensation hypothesis for the Antarctic Circumpolar Current in the Southern Ocean, where eddies appear to dampen the effect of increased wind energy input on the mean flow³⁰. In essence, while winds tend to accelerate the flow and steepen isopycnals, eddies mix laterally across the current to slump the isopycnals. Coupling between eddies and the atmosphere has also been shown to influence this frontal balance³¹. The implication of broadening boundary currents is a more porous divide between the continental shelves and the open ocean, leading to greater mixing and cross-frontal exchange. In the Agulhas Current, these changes could also enhance upwelling over the shelf, since the strongest upwelling events are driven by meanders. These implications are in contrast to those of an intensifying flow, which would tend to dampen cross-frontal mixing and increase meridional heat transport.

If western boundary currents are not strengthening, observed patterns of surface warming^{5,9,10} must be explained by a poleward shift of the ocean gyres. Broadening of western boundary currents and their extensions will also imprint on warming patterns and this should be considered in future analyses. Ocean reanalysis products and climate models fail to resolve western boundary currents and this could explain discrepancies among them and with our results^{5,10}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 May; accepted 9 September 2016.

Published online 9 November; corrected online 21 December 2016
(see full-text HTML version for details).

- Hu, D. *et al.* Pacific western boundary currents and their roles in climate. *Nature* **522**, 299–308 (2015).
- Palter, J. B. The role of the Gulf Stream in European climate. *Annu. Rev. Mar. Sci.* **7**, 113–137 (2015).
- Cai, W. Antarctic ozone depletion causes an intensification of the Southern Ocean super-gyre circulation. *Geophys. Res. Lett.* **33**, L03712 (2006).
- McGregor, S., Gupta, A. S. & England, M. H. Constraining wind stress products with sea surface height observations and implications for Pacific Ocean sea level trend attribution. *J. Clim.* **25**, 8164–8176 (2012).
- Yang, H. *et al.* Intensification and poleward shift of subtropical western boundary currents in a warming climate. *J. Geophys. Res. Oceans* **121**, 4928–4945 (2016).
- Sverdrup, H. U. Wind-driven currents in a baroclinic ocean; with application to the equatorial currents of the eastern Pacific. *Proc. Natl Acad. Sci. USA* **33**, 318–326 (1947).
- Stommel, H. The westward intensification of wind-driven ocean currents. *Eos* **29**, 202–206 (1948).

- Saenko, O. A., Fyfe, J. C. & England, M. H. On the response of the oceanic wind-driven circulation to atmospheric CO₂ increase. *Clim. Dyn.* **25**, 415–426 (2005).
- Rouault, M., Penven, P. & Pohl, B. Warming in the Agulhas Current system since the 1980's. *Geophys. Res. Lett.* **36**, L12602 (2009).
- Wu, L. *et al.* Enhanced warming over the global subtropical western boundary currents. *Nat. Clim. Chang.* **2**, 161–166 (2012).
- Uchida, H. & Imawaki, S. Estimation of the sea level trend south of Japan by combining satellite altimeter data with *in situ* hydrographic data. *J. Geophys. Res. Oceans* **113**, C09035 (2008).
- Oliver, E. & Holbrook, N. Extending our understanding of South Pacific gyre spin-up: modeling the East Australian Current in a future climate. *J. Geophys. Res. Oceans* **119**, 2788–2805 (2014).
- Cetina-Heredia, P., Roughan, M., Van Sebille, E. & Coleman, M. Long-term trends in the East Australian Current separation latitude and eddy driven transport. *J. Geophys. Res. Oceans* **119**, 4351–4366 (2014).
- Yan, X. & Sun, C. An altimetric transport index for Kuroshio inflow northeast of Taiwan Island. *Sci. China Earth Sci.* **58**, 697–706 (2015).
- Sloyan, B. M. & O'Kane, T. J. Drivers of decadal variability in the Tasman Sea. *J. Geophys. Res. Oceans* **120**, 3193–3210 (2015).
- Beal, L. M., Elipot, S., Houk, A. & Leber, G. M. Capturing the transport variability of a western boundary jet: results from the Agulhas Current Time-Series experiment (ACT)*. *J. Phys. Oceanogr.* **45**, 1302–1324 (2015).
- Elipot, S. & Beal, L. M. Characteristics, energetics, and origins of Agulhas Current meanders and their limited influence on ring shedding. *J. Phys. Oceanogr.* **45**, 2294–2314 (2015).
- Tsugawa, M. & Hasumi, H. Generation and growth mechanism of the Natal Pulse. *J. Phys. Oceanogr.* **40**, 1597–1612 (2010).
- Sprintall, J. & Révelard, A. The Indonesian throughflow response to Indo-Pacific climate variability. *J. Geophys. Res. Oceans* **119**, 1161–1175 (2014).
- Backeberg, B. C., Penven, P. & Rouault, M. Impact of intensified Indian Ocean winds on mesoscale variability in the Agulhas system. *Nat. Clim. Chang.* **2**, 608–612 (2012).
- Compo, G. P. *et al.* The twentieth century reanalysis project. *Q. J. R. Meteorol. Soc.* **137**, 1–28 (2011).
- Munk, W. H. On the wind driven ocean circulation. *J. Meteorol.* **7**, 80–93 (1950).
- Bryden, H. L., Beal, L. M. & Duncan, L. M. Structure and transport of the Agulhas Current and its temporal variability. *J. Oceanogr.* **61**, 479–492 (2005).
- Pedlosky, J. *Geophysical Fluid Dynamics* 2nd edn (Springer, 1987).
- England, M. H. *et al.* Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nat. Clim. Chang.* **4**, 222–227 (2014).
- Cheng, W., Chiang, J. C. & Zhang, D. Atlantic meridional overturning circulation (AMOC) in CMIP5 models: RCP and historical simulations. *J. Clim.* **26**, 7187–7197 (2013).
- Ezer, T., Atkinson, L. P., Corlett, W. B. & Blanco, J. L. Gulf Stream's induced sea level rise and variability along the US mid-Atlantic coast. *J. Geophys. Res. Oceans* **118**, 685–697 (2013).
- Rosby, T., Flagg, C., Donohue, K., Sanchez-Franks, A. & Lillibridge, J. On the long-term stability of Gulf Stream transport based on 20 years of direct measurements. *Geophys. Res. Lett.* **41**, 114–120 (2014).
- Kent, E. C., Fangohr, S. & Berry, D. I. A comparative assessment of monthly mean wind speed products over the global ocean. *Int. J. Climatol.* **33**, 2520–2541 (2013).
- Farneti, R., Delworth, T. L., Rosati, A. J., Griffies, S. M. & Zeng, F. The role of mesoscale eddies in the rectification of the Southern Ocean response to climate change. *J. Phys. Oceanogr.* **40**, 1539–1557 (2010).
- Ma, X. *et al.* Western boundary currents regulated by interaction between ocean eddies and the atmosphere. *Nature* **535**, 533–537 (2016).

Acknowledgements L.M.B. and S.E. acknowledge support from the US National Science Foundation, grant OCE-0850891. The altimeter products distributed by Aviso were produced by Ssalto/Duacs, with support from Cnes (<http://www.aviso.altimetry.fr/duacs/>).

Author Contributions L.M.B. designed, obtained funding for, and led the Agulhas Current Time-series experiment and wrote and edited this paper. S.E. performed the analyses and wrote the Methods section.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.M.B. (lbeal@rsmas.miami.edu).

METHODS

Mooring array and transports. Our *in situ* data are from the ACT mooring array, deployed across the Agulhas Current and along a satellite altimeter ground track near 34° S between April 2010 and February 2013 (Fig. 1). The array is 300 km long and oriented 15° clockwise from normal to the mean flow. Seven full-depth current-meter moorings and four current- and pressure-sensor-equipped inverted echo sounders (CPIES) were able to capture the full Agulhas jet at all times, including during meander events¹⁶. CPIES pairs provide a cost-effective estimate of full-depth geostrophic flow at the offshore end of the array. Cross-track velocity profiles are horizontally interpolated to a resolution of 1 km and integrated vertically to obtain transport per unit distance, T_x . We then define the western boundary jet transport T_{jet} as the southwestward component of T_x integrated to the first maximum of T_x beyond the half-width of the mean jet (110 km)¹⁶. We define boundary layer transport T_{box} as T_x integrated out to 219 km, the three-year mean width of the jet (Fig. 1).

Absolute dynamic topography. We use the 1-Hz unfiltered along-track absolute dynamic topography (ADT) product from Aviso. ADT is the sum of a mean dynamic topography and a sea level anomaly. The latest Aviso product uses a 20-year reference period from 1993 to 2012 to define the mean dynamic topography. Using sea level anomaly instead of ADT does not change our statistical analyses and regressions, since the two quantities differ only by a constant. In the main text we refer to ADT as sea surface height or sea level, although strictly the definition of ADT differs from sea surface height by a constant geoid. The ACT line is along altimeter track number 96 successively occupied by satellites TOPEX/Poseidon (1992–2002), Jason-1 (2002–2008), and currently Jason-2 (since 2008). During the ACT experiment, there were 105 satellite passes across the Agulhas Current, providing data between 14.6 km and 306 km offshore with a horizontal resolution of about 6.2 km. Data are available up to 8.45 km from the coast, but we discount them because they are missing 30% of the time. We estimate ADT and its slope every 1 km along the track using a local order-one polynomial regression estimator with a 24-km half-bandwidth and an Epanechnikov kernel³². Estimating slope in this way introduces less noise than differentiation and the Epanechnikov kernel minimizes the asymptotic mean square error of the resulting estimates³². A half-bandwidth of 24 km corresponds to a total window length that approaches the horizontal along-track decorrelation length scale of the flow at the ACT array (56 km)¹⁶. Varying this bandwidth by 50% does not significantly modify our results.

Transport proxies. To test the assumption of a linear and fixed relationship between *in situ* transport and sea surface height and slope for the Agulhas Current we compare the combined eigenmodes of variance of ADT and ADT slope over the full 22-year record to those of T_x over our 3 years of measurements. In each case we find four eigenmodes, of similar spatial structure, that each explain 10% or more of the variance (Extended Data Fig. 1). The dominant mode of variance at the sea surface is a broad-scale decrease or increase in sea surface slope, which is associated with a weakening or strengthening of the *in situ* transport. This is demonstrated by a strong correlation between the Principal Component time series of the first ADT mode and that of the second mode of T_x (correlation 0.76). A narrowing or broadening of the jet is reflected in the second mode of ADT and the first mode of T_x (correlation 0.84). The remaining third and fourth modes of variance in each case reflect meandering of the jet (correlations 0.72 and 0.71, respectively), while all four eigenmodes project onto mesoscale meander events. In all cases we find P values smaller than 10^{-3} and hence all correlations are highly significant. Because the Principal Component time series of T_x are serially correlated and not normally distributed, we used a nonparametric resampling method³³ to calculate one-tailed P values for the magnitude of correlation. Given the similarities between eigenmodes and the significance of correlations, we conclude that a linear relationship between sea surface variance and full-depth Agulhas transport is a fair assumption at 34° S, where the undercurrent is weak¹⁶. Hence, the variance of the current appears equivalent barotropic, in agreement with previous analyses^{17,18}.

For our preferred proxy we build nine regression models, one at each current-meter mooring and CPIES pair along the ACT array, which linearly relate the local T_x to the slope of ADT. Estimating ADT slope requires careful determination of the horizontal length scale of the flow at each site. We achieve this by using again an order-one local polynomial regression estimator with an Epanechnikov kernel³² to estimate both ADT and its slope, but this time varying the spatial bandwidth of the estimator in order to maximize the correlation between our measured T_x and the slope estimate at each mooring. This gives length scales ranging from 27 km at mooring B to 102 km at mooring G, consistent with our physical expectation of increasing length scale with increasing distance offshore. The polynomial regression estimator allows us to calculate the error variance of ADT and slope estimates at each mooring, based on the measurement errors in along-track ADT reported by Aviso. Reassuringly, we find that these error variances are at least an order of magnitude smaller than the variance of the resulting ADT and slope time series. Next, we remove outlying slope estimates by

discarding the upper and lower 0.25% of the data distribution. To fill gaps we use multivariate regression between ADT slope at the location with missing data and at surrounding locations. This technique recovers 90% of the variance of the missing data, except at mooring A, where only 68% of the variance can be explained using adjacent records. Finally, we build linear regression models between ADT slope and T_x at each of the nine mooring sites. The R^2 statistics of these models vary from 0.51 at mooring A to 0.81 at CPIES pair P4 and P5. The strongly sloping sea bed and Agulhas undercurrent (Fig. 1) probably contribute to the poorer skill of the regressions at the inshore moorings.

From the results of the regressions, we obtain T_x at each of the nine mooring locations for each satellite pass over 22 years. Subsequently, to calculate total transports from these discrete points, we fit a shape-preserving piecewise cubic Hermite interpolating polynomial function to obtain T_x at 1-km intervals across the current. We then integrate to obtain T_{jet} and T_{box} for each altimeter pass, as defined above (Fig. 2). Using instead a piecewise linear interpolation of T_x between moorings and applying a low-pass filter with a cutoff of 56 km (the decorrelation length scale given by the ACT measurements) to produce a smooth function gives alternate estimates with root-mean-square differences of 4.6 Sv for T_{jet} and 1.3 Sv for T_{box} .

A more common methodology is to build a proxy by regressing total transport onto a broad-scale sea surface slope^{14,19}. Regression based directly on along-track ADT at the ACT array is problematic, because the sea surface is strongly covarying along the length of the array, leading to large uncertainties in the regression coefficients (even though the resulting R^2 statistics of such a proxy can be high). Hence, we build a regression model using the uncorrelated Principal Components of the combined EOFs of ADT and ADT slope from above (Extended Data Fig. 1). To include as many altimeter passes as possible for the regression model, we recompute these combined EOFs for data points farther than 42.6 km from the shore to avoid missing data. The resulting EOFs differ very little from the ones presented in Extended Data Fig. 1. The regression model for both the jet and boundary transport proxies T is then:

$$T = \alpha_0 + \sum_{k \in \Omega} \alpha_k A_k(t) \quad (1)$$

where α_k are the regression coefficients estimated using the ordinary least-squares method and Ω is an ensemble of the Principal Component time series $A_k(t)$. For the ensemble we consider each eigenmode, in order of decreasing amount of variance explained, and incorporate successively only those Principal Components that increase the adjusted- R^2 statistics of our model (Equation (1)). The adjusted- R^2 statistics quantifies the amount of variance explained by a multivariate model but, in contrast to the classic R^2 statistics, attempts to correct for the explained variance that occurs by chance from adding a random variate to the model. For T_{box} this procedure selects Principal Components 1 and 2 (correlated with transport at 0.85 and 0.22 respectively, with P values $< 10^{-3}$ and 0.09), and for T_{jet} Principal Components 1, 2 and 3 (correlated at 0.49, 0.27 and 0.20, respectively, with P values $< 10^{-3}$, 0.015 and 0.05).

The resulting proxy for T_{box} explains 76% of the variance, while in the case of T_{jet} the proxy explains only 39% of the variance (Extended Data Fig. 2). The poor performance of the jet proxy is probably because it is defined from only the southwestward flow, while sea level variance along the whole array reflects the net flow. Estimating the linear trends of the two time series we find that the boundary layer transport appears to be strengthening at a significant rate ($\pm 95\%$ confidence interval) of -4.9 ± 2.0 Sv per decade, while the jet transport has a weaker and insignificant trend of -1.4 ± 1.5 Sv per decade (Extended Data Fig. 2). The boundary layer proxy trend fulfils prior expectations, based on warming sea surface temperatures and strengthening winds over the region^{9,10}. However, this strengthening is inconsistent with trends in sea surface height across the current (Fig. 3). The structures and nodes of the eigenmodes of sea level and slope (Extended Data Fig. 1) are fixed in space and seem to be unable to capture the change in structure of the boundary layer over time. Throughout the 22-year record, the trend patterns explain never more than 8% of the local variance and the EOFs are virtually unchanged when recomputed after detrending the data. Simply put, a proxy based on a single regression with total transport does not allow for a broadening of the current.

Spectral estimates and annual cycle. Power spectral densities of the proxies T_{jet} and T_{box} are estimated using the adaptive multitaper method with five Slepian tapers³⁴. This method exhibits less variance than other common methods and is routinely employed for climate time series analysis³⁵. Using three or seven tapers, of the Slepian form or other forms, does not qualitatively change our results. Significant spectral peaks at the 95% confidence level were identified using a test for periodicities within a red background spectrum³⁶. Since T_{jet} has less energy than T_{box} at periods longer than 100 days (Fig. 2), we expect decadal signals and trends to be more clearly detected in the jet transport. Annual cycles are obtained by calculating daily means using the Nadaraya-Watson kernel estimator with a

Gaussian kernel of half-width 30 days³². 95% confidence intervals are based on the assumption of a normal distribution of the estimates.

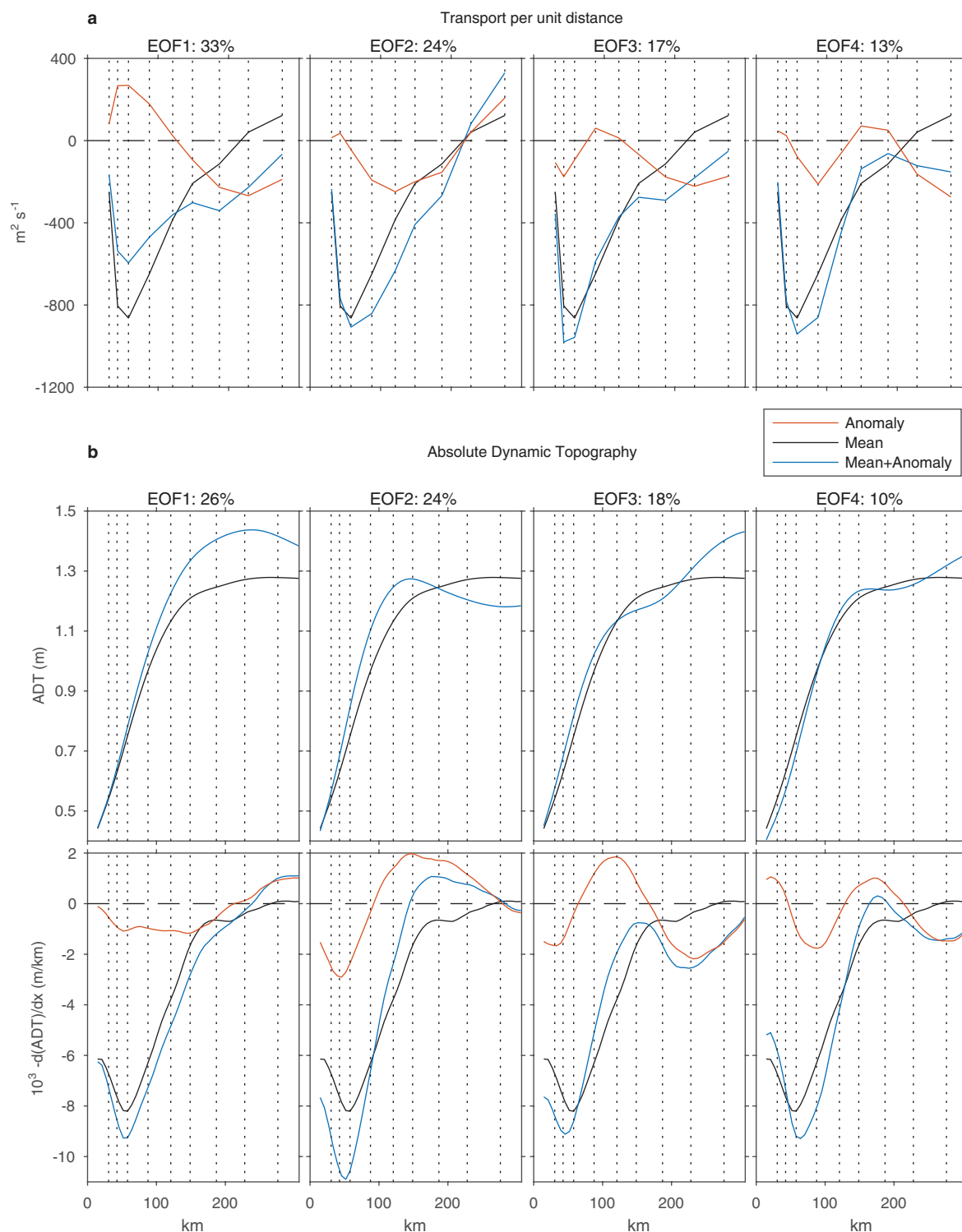
Trends in sea level and kinetic energy. Trends in sea level (Fig. 3) are calculated using along-track satellite altimeter data from Aviso and from the new Sea Level Climate Change Initiative (CCI) product³⁷, which is optimized for long-term signals. For kinetic energy across the Agulhas Current (Fig. 4) we use the post-2014 geostrophic velocity maps from Aviso, which are derived from a two-satellite, merged, delayed-time product. For long-term signals it is important to use the two-satellite product because it ingests a consistent amount of data over time, avoiding the introduction of sampling bias that may cause spurious trends in variance. TKE is defined as half the sum of the squared horizontal components of velocity. MKE is calculated from the total velocity time series by low-pass filtering using a sliding quadratic window with half-bandwidth of 18 months. EKE is calculated from the velocity residuals obtained by subtracting the low-pass-filtered velocities from the total velocities. These calculations result in an interannually evolving MKE, and an EKE which captures variability at timescales of less than 18 months. Although mapped satellite altimeter products have difficulties resolving boundary flows that are narrow and close to the coast, reassuringly we find that at the ACT array peak MKE occurs close to the mean *in situ* Agulhas Current core (Fig. 4a).

Code availability. MATLAB scripts used for our analyses and figures are available upon request from S.E.

Data availability. The *in situ* mooring data from the ACT experiment (velocities from current meters and acoustic Doppler current profilers; sound speed, pressure

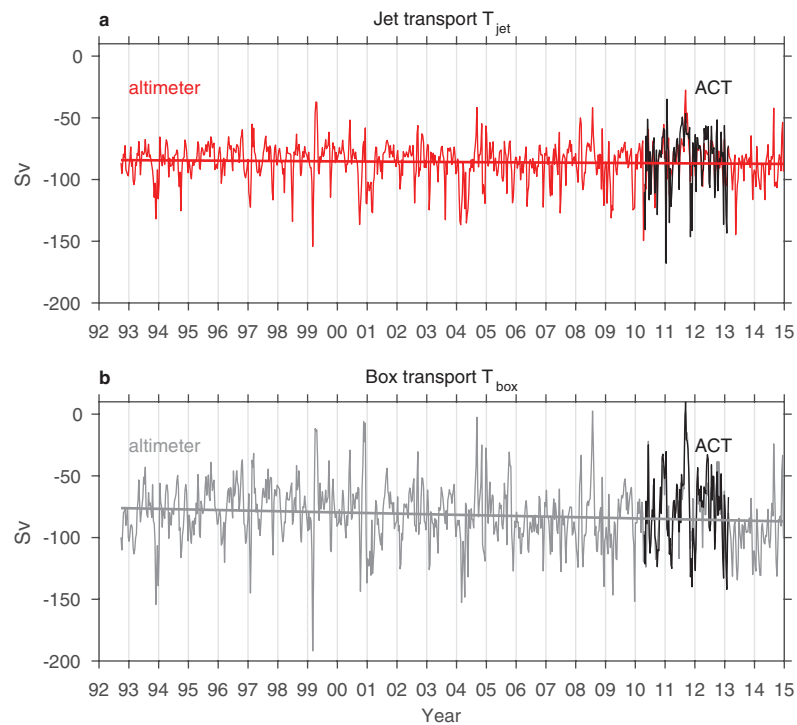
and bottom velocity from CPIES) are archived with the NOAA National Centers for Environmental Information (<https://www.ncei.noaa.gov>), with accession numbers 0156669 and 0156605. Aviso along-track Absolute Dynamic Topography data are accessible through <http://www.aviso.altimetry.fr/en/data/products/sea-surface-height-products/global/adt-h.html#c5139>. Maps of absolute dynamic topography and absolute geostrophic velocities are accessible through <http://www.aviso.altimetry.fr/en/data/products/sea-surface-height-products/global/madt-h-uv.html>. The CCI altimeter along-track data correspond to the Fundamental Climate Data Record product (http://dx.doi.org/10.5270/esa-sea_level_cci-1993_2014-v1.1-201512) generated by the Sea Level Climate Change Initiative project (<http://www.esa-sealevel-cci.org/>).

32. Fan, J. & Gijbels, I. *Local Polynomial Modelling and its Applications: Monographs on Statistics and Applied Probability* Vol. 66 (CRC Press, 1996).
33. Ebisuzaki, W. A method to estimate the statistical significance of a correlation when the data are serially correlated. *J. Clim.* **10**, 2147–2153 (1997).
34. Thomson, D. Spectrum estimation and harmonic analysis. *Proc. IEEE* **70**, 1055–1096 (1982).
35. Grimm, E. C. *et al.* A 50,000-year record of climate oscillations from Florida and its temporal correlation with the Heinrich events. *Science* **261**, 198 (1993).
36. Percival, D. B. & Walden, A. T. *Spectral Analysis for Physical Applications* (Cambridge Univ. Press, 1993).
37. Ablain, M. *et al.* Improved sea level record over the satellite altimetry era (1993–2010) from the climate change initiative project. *Ocean Sci.* **11**, 67–82 (2015).



Extended Data Figure 1 | Spatiotemporal variance of oceanic transport and sea surface slope are similar across the Agulhas Current. **a**, First four eigenmodes of the transport per unit distance from the ACT array. **b**, First four combined eigenmodes of sea surface height and slope, from

Aviso along-track absolute dynamic topography (ADT) from satellite altimetry. Note that the ADT gradient is positive across most of the array, but is shown as negative for comparison with the southward Agulhas Current transports in **a**. Black dotted lines depict mooring positions.



Extended Data Figure 2 | Agulhas Current transport proxies based on regressions of total transport with sea surface eigenmodes. a, Proxy for jet or stream-wise transport T_{jet} . **b,** Proxy for boundary layer transport T_{box} . The three years of *in situ* transports from the ACT array are shown as black lines. The trends of these two proxies are inconsistent, owing to observed broadening of the jet.

Chemical intervention in plant sugar signalling increases yield and resilience

Cara A. Griffiths^{1*}, Ram Sagar^{2*†}, Yiqun Geng^{2*}, Lucia F. Primavesi¹, Mitul K. Patel², Melissa K. Passarelli³, Ian S. Gilmore³, Rory T. Steven³, Josephine Bunch^{3,4}, Matthew J. Paul¹ & Benjamin G. Davis²

The pressing global issue of food insecurity due to population growth, diminishing land and variable climate can only be addressed in agriculture by improving both maximum crop yield potential and resilience^{1,2}. Genetic modification is one potential solution, but has yet to achieve worldwide acceptance, particularly for crops such as wheat³. Trehalose-6-phosphate (T6P), a central sugar signal in plants, regulates sucrose use and allocation, underpinning crop growth and development^{4,5}. Here we show that application of a chemical intervention strategy directly modulates T6P levels *in planta*. Plant-permeable analogues of T6P were designed and constructed based on a ‘signalling-precursor’ concept for permeability, ready uptake and sunlight-triggered release of T6P *in planta*. We show that chemical intervention in a potent sugar signal increases grain yield, whereas application to vegetative tissue improves recovery and resurrection from drought. This technology offers a means to combine increases in yield with crop stress resilience. Given the generality of the T6P pathway in plants and other small-molecule signals in biology, these studies suggest that suitable synthetic exogenous small-molecule signal precursors can be used to directly enhance plant performance and perhaps other organism function.

We designed a signalling-precursor strategy on the basis of release by light (Extended Data Fig. 1). Light-activated control is a potent modulation strategy in biology, allowing for temporal and spatial resolution surpassing that of standard genetic methods⁶. Additionally, such resolution can be increased when combined with small-molecule chemical control^{7–9}. Potency is further increased when releasing a signalling molecule, as the effects of light-activated control are increased several-fold through the inherent amplification of signalling.

Hydrophilic or charged molecules do not readily enter plants unless transported. We therefore designed unnatural precursors (**1–4**) of T6P with groups to mask charge, increase hydrophobicity and also to induce release by light (Fig. 1a). Their construction (Fig. 1b) used different phosphorus chemistries: phosphoramidite chemistry^{10,11} to create P(III)-intermediates that were then oxidized to corresponding P(V)-phosphotriesters, or direct P(V)-phosphorylation chemistry (Fig. 1b). Regioselective access to the OH-6 group in trehalose exploited trimethylsilyl as a protecting group that is chemically orthogonal to the phosphotriester; **12** was prepared on a multigram scale¹². Phosphorylation (reaction with phosphoramidites **9–11** (refs 10, 11) followed by tBuOOH, or treatment with POCl₃ (ref. 13) followed by the addition of the appropriate alcohol) gave intermediates that were deprotected under mildly acidic conditions (see Supplementary Methods). **1–4** were all inactive against SnRK1 (SNF1-related kinase 1) (Extended Data Fig. 2).

Mass spectrometry, thin-layer chromatography and nuclear magnetic resonance spectroscopy (see Supplementary Methods, Supplementary Table 1 and Extended Data Fig. 1) showed release times (95% release, *t*₉₅)

dependent on both light intensity and frequency under a range of conditions. Consistent with design, light-sensitive groups were differently susceptible. Precursor *ortho*-nitrobenzyl (*o*NB)-T6P **1**, for example, generated T6P by light-activated release more rapidly at lower wavelengths, while compound mono-dimethoxy(*ortho*-nitro)benzyl (mono-DMNB)-T6P **4** was more reactive at higher wavelengths. Although release with higher light intensity/photon flux (125 W/365 μmol m⁻² s⁻¹ (with photon flux defined as the number of photons per m² per second) compared to 8 W/23 μmol m⁻² s⁻¹) was more rapid, direct sunlight proved sufficient, in some cases resulting in *t*₉₅ as brief as 90 min (for (*ortho*-nitrophenyl)ethyl (*o*NPE)-T6P (compound **3**)). Nuclear magnetic resonance spectroscopy analysis (Supplementary Methods and Extended Data Fig. 1e, f) confirmed that T6P was formed and that potent inhibitory activity against SnRK1 was induced (Extended Data Fig. 2).

Following successful *in vitro* release, uptake *in planta* was examined. Compounds **1–4** (at a final concentration of 1 mM) were fed to roots of plantlets of *Arabidopsis thaliana* and the aerial parts were analysed over time and with increasing dose (Fig. 2 and Supplementary Tables 2–7). High-performance liquid chromatography (HPLC) followed by quantitative mass spectrometry¹⁴ (HPLC–MS) of extracts of the aboveground biomass (shoot and leaves) showed increasing uptake of the compounds over time (Fig. 2a and Extended Data Fig. 3) and with increasing dose. Consistent with design, structural variation showed that altered hydrophobicity modulated permeability¹⁵ and transport¹⁶. Notably, systematic variation of group type and copy number identified compound **3** (log[*P*] of the compound is 0.11 ± 0.60, where *P* is the partition coefficient; see also Supplementary Methods and Supplementary Table 17) as the most readily taken up (Fig. 2a), with absorption of approximately 20% after 72 h. Compounds **1**, **2** (dimethoxy(*ortho*-nitro)benzyl (DMNB)-T6P) and **4** (log[*P*] of the compounds ranged from –2.35 to –0.17), on the other hand, were less readily taken up.

Next, we investigated the light-activated release *in planta*. Plants were treated with compounds dissolved in medium, grown for a further three days, irradiated, and the shoots were harvested and extracted¹⁷. T6P release was confirmed by tandem mass spectrometry (Extended Data Fig. 4) and determined by quantitative HPLC–MS (with 2-deoxyglucose-6-phosphate as an internal standard¹⁴; Fig. 2b, Extended Data Fig. 4 and Supplementary Table 10). Release *in planta* could be controlled and modulated by the choice of light source and signalling precursor (Fig. 2b and Supplementary Table 10).

Most transgenic approaches only alter T6P over a 2–3-fold range^{4,5}. Using compounds **1–4**, levels of up to 900 nmol g⁻¹ fresh weight were attainable, which was 100-fold higher than endogenous levels and 75-fold higher than can be achieved with genetic methods. Consistent with this strategy, maximal T6P was released when precursor-treated plants were irradiated with the highest flux (100 W/292 μmol m⁻² s⁻¹ UV) in all cases. Notably, and with relevance to application in the field,

¹Plant Biology and Crop Science, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK. ²Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford OX1 3TA, UK. ³National Centre of Excellence in Mass Spectrometry Imaging (NiCE-MSI), National Physical Laboratory (NPL), Teddington, Middlesex TW11 0LW, UK. ⁴School of Pharmacy, University of Nottingham, Nottingham NG7 2RD, UK. [†]Present address: Department of Chemistry, School of Natural Sciences, Shiv Nadar University, Greater Noida 201 314, India.

*These authors contributed equally to this work.

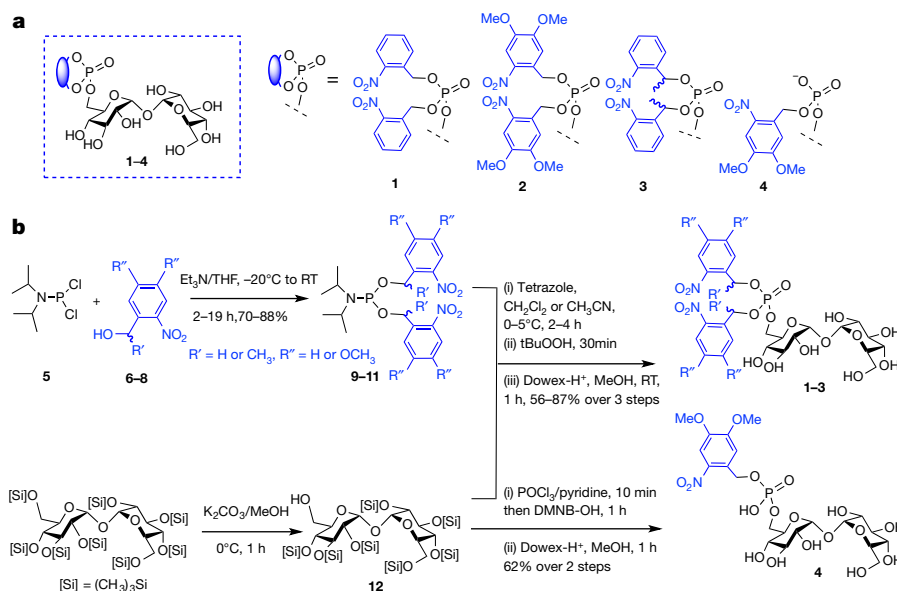


Figure 1 | Design and synthesis of signalling precursors of T6P. T6P is plant-impermeable; synthesis of plant-permeable variants allowed subsequent photo-activated release of T6P *in planta*. **a**, Designed precursors 1–4, with the backbone on the left and side chains (blue circle on the back bone) of the individual precursors on the right. **b**, Synthesis

with sunlight only, all treated plants released significantly increased amounts of T6P ($39\text{--}296\text{ nmol g}^{-1}$ fresh weight) (Supplementary Table 10), some approximately 4–30-fold above endogenous levels. There was no significant reduction in the fresh weight of plantlets treated with 1 mM of compound (Supplementary Methods and Extended Data Fig. 5), suggesting low toxicity. Accumulation of T6P after treatment was analysed with mass spectrometry imaging¹⁸ (Fig. 3) using signature-ion markers¹⁹ in treated leaves of *A. thaliana* seedlings (Fig. 3b) after 2 h of irradiation; the different distributions from compounds 2 and 3 appeared consistent with their measured release rates. Notably, increased trehalose was also observed²⁰ in the same regions (Fig. 3c), suggesting metabolism. Moreover, mass spectrometry imaging using treatment-specific ions corroborated the uptake of precursors into leaves (Fig. 3d, e).

The dynamics of this enhanced *in planta* T6P release, and possible consequent metabolic products, were determined not only through both quantitative HPLC–MS and/or enzymatic quantification, but

of the precursors using phosphoramidite chemistry (1–3) or direct phosphorylation chemistry (4) from the key intermediate compound 12. Universally ^{13}C -labelled 2* was prepared in essentially the same manner (see Extended Data Fig. 7a). THF, tetrahydrofuran.

also through the use of unnaturally enriched isotopic labelling of the signalling precursors, allowing for unambiguous delineation of their fate (Fig. 4, Extended Data Fig. 6 and Supplementary Methods). Thus, in 7-day-old *A. thaliana* seedlings²¹, treatment with 1 mM of compound 2 or 3, fed for 24 h before exposure to light (UV 8 W/23 $\mu\text{mol m}^{-2}\text{ s}^{-1}$), led to peak T6P after 60 min (229 and 159 nmol g^{-1} fresh weight, respectively), which declined over the following 2 days (Fig. 4a). Corresponding trehalose levels were also elevated, with peaks at around 2 h (up to a maximum of 134 nmol g^{-1} fresh weight compared to control levels of 20 nmol g^{-1} fresh weight; Fig. 4b), confirming the metabolism indicated by mass spectrometry imaging. Glucose, the next sugar in the pathway, was also increased but to a smaller degree, peaking at around 2–4 h (Fig. 4d). These levels are consistent with known low metabolic fluxes²². Given known interrelationships⁵, sucrose levels were also determined. Notably, these increased 2–3-fold over the first 2 h of irradiation (Fig. 4c) and correlated positively with T6P for both compounds 2 and 3 (Fig. 4); whereas fructose was minimally affected

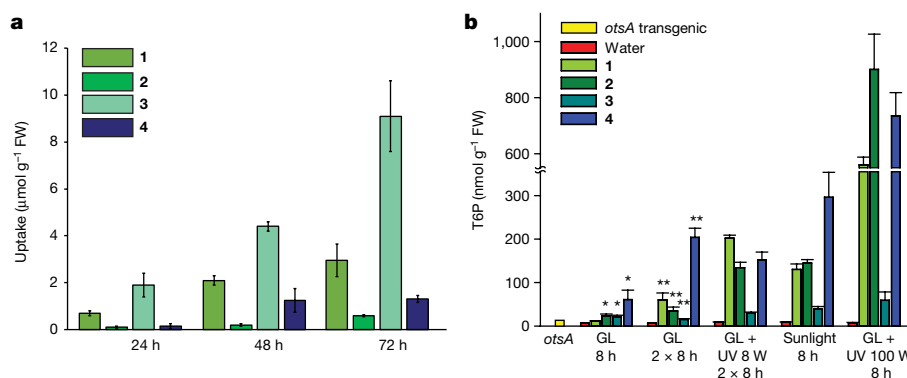


Figure 2 | In planta uptake of signalling precursors and T6P release. **a**, Uptake of compounds 1–4 (1 mM in medium) at 24, 48 and 72 h (data are shown as mean \pm s.e.m., $n = 3$). FW, fresh weight. **b**, T6P released *in planta* (data are shown as mean \pm s.e.m., $n = 3$). Compounds were applied 18 days after sowing and then irradiated for 72 h. *otsA*, *A. thaliana* overexpressing trehalose-phosphate synthase. Growth light (GL) and other conditions are defined as the following levels of photon flux: GL, 250 $\mu\text{mol m}^{-2}\text{ s}^{-1}$; GL + 8 W UV (365 nm), 23 $\mu\text{mol m}^{-2}\text{ s}^{-1}$; GL + 100 W UV,

292 $\mu\text{mol m}^{-2}\text{ s}^{-1}$; sunlight (cloud), 250 $\mu\text{mol m}^{-2}\text{ s}^{-1}$; and sunlight (no cloud), 1,440 $\mu\text{mol m}^{-2}\text{ s}^{-1}$. ANOVA showed significant differences ($P < 0.001$) between treatments (water or precursor) for each regime. All treatments with precursor and UV showed significance ($P < 0.001$, least significant difference (LSD)) compared to water and UV. For growth light irradiance * $P < 0.05$; ** $P < 0.01$ (LSD, data shown on a linear scale). See also Supplementary Table 10.

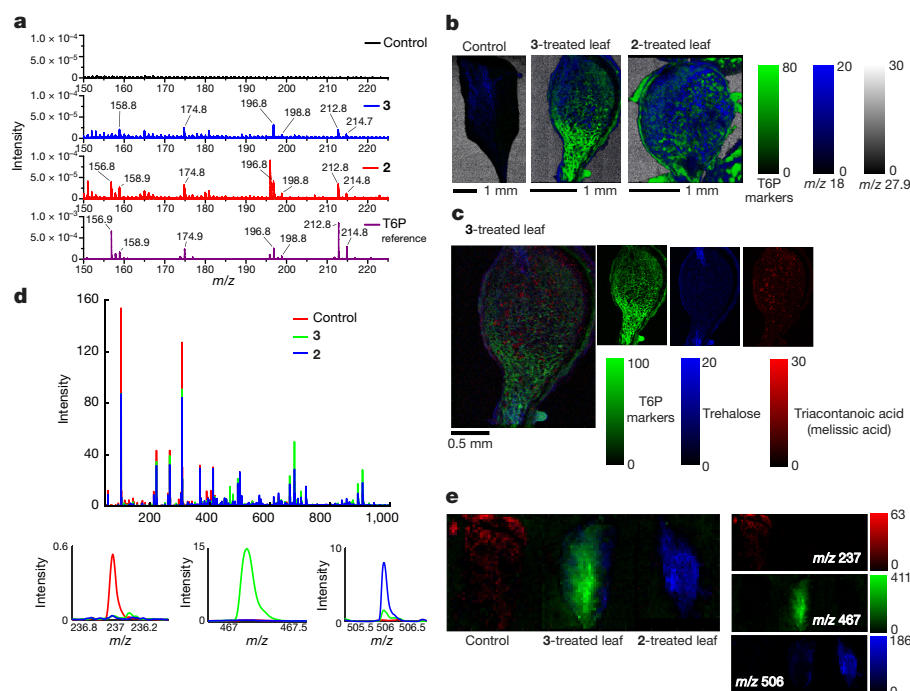


Figure 3 | Mass spectrometry imaging of treated leaves. For all ion images, the pixel intensity scale represents the area under the corresponding m/z peak. **a**, Time-of-flight secondary ion mass spectrometry (ToF-SIMS) spectra from the surface of *A. thaliana* leaves (top three spectra). Marker ions m/z 156.8, 196.8 and 212.8 in T6P reference (bottom spectrum). **b**, T6P (three markers, green) in control (left), 3-treated (middle) and 2-treated (right) leaves, anti-colocalized with H_2O (m/z 18.0) and silicon substrate (m/z 27.9). **c**, T6P (markers,

green, positive mode), trehalose (m/z 325.2, 321.2, blue, negative mode) and overlay with known¹⁹ epicuticular wax markers (red, negative mode) in the 3-treated leaf. **d**, **e**, Matrix-assisted laser desorption/ionization-mass spectrometry (MALDI-MS). **d**, Overlay of mean on-leaf spectra for control, 3-treated and 2-treated leaves. Lower panels show expansions for correlated markers. **e**, RGB-colour overlay images of marker ions; separate ion images are shown on right. Images in **b**, **c** and **e** are representative of three individual images.

(Extended Data Fig. 6d). We confirmed that inhibited growth was not an explanation for sucrose accumulation and found instead that growth was stimulated by T6P (Extended Data Fig. 6e).

Creation of a ^{13}C -isotopically-labelled variant 2* (Extended Data Fig. 7a) allowed direct tracking via 'mass-shifts' of the corresponding ions using mass spectrometry. Treatment with 2* led to release of ^{13}C -T6P and consequent sequential metabolism (to ^{13}C -trehalose and ^{13}C -glucose) following essentially the same dynamics (Extended Data Fig. 7). Notably, using this method, mass labelling also showed that the compounds not only released, but also induced T6P. This accounted for approximately half of T6P measured at 30 min, thereby providing direct evidence of induction of *de novo* T6P synthesis and this induction continued, giving rise to increased T6P accumulation over time (Extended Data Fig. 7). This could be due to the large increase in sucrose observed,

as sucrose induces T6P^{5,21}. Together these data suggest that perturbation of T6P levels occurs via two modes of action: direct release from the signalling precursor and simultaneously induced biosynthesis of T6P by the plant.

Our data indicate that compound 3 is the plant-permeable signalling precursor with the greatest tissue uptake coupled with the greatest temporal control (consistent with its tuned permeability and its fastest release rates), allowing minimal application amounts (0.1 mM), while still able to enhance T6P levels around 1.5–6.5-fold above endogenous levels without potentially detrimental disruption of metabolism²³. Plants were treated with precursor 3 for 72 h and were then subjected to a single 8-h period under growth lights supplemented with 8 W UV (Supplementary Table 10, generating around 21 nmol g^{-1} fresh weight T6P) and harvested a day later.

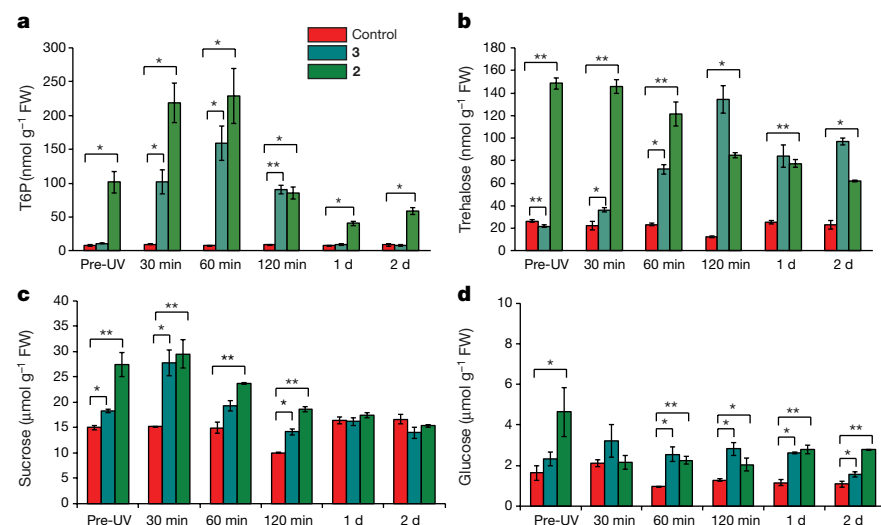


Figure 4 | In planta T6P release and sugar metabolism over time. **a–d**, Seven-day-old *A. thaliana* seedlings grown in liquid culture were treated with 1 mM of either 3 or 2; control seedlings were treated with water. Seedlings were left under growth lights to take up the signalling precursors for 24 h, after which the plants were exposed to 23 μ mol $m^{-2} s^{-1}$ UV for 2 h. Measurements were taken 1 d after uptake (pre-UV), 30, 60 and 120 min after the initiation of UV treatment (23 μ mol $m^{-2} s^{-1}$), and 1 and 2 days after the initiation of UV treatment. **a**, T6P content. **b**, Trehalose content. **c**, Sucrose content. **d**, Glucose content. In all cases $n = 3$; * $P < 0.05$; ** $P < 0.01$; Student's *t*-test; data are shown as mean \pm s.e.m.

The mean starch level ($63.2 \mu\text{mol g}^{-1}$ fresh weight) determined²⁴ was significantly higher ($F_{(1,14)} = 13.59$; $P = 0.002$) than for water-treated plants ($40.7 \mu\text{mol g}^{-1}$ fresh weight; Extended Data Fig. 5d).

The production of T6P from compounds 1–4 involves fragmentation with concomitant release of side products. Although considered²⁵ to be non-toxic, we nevertheless tested for any unexpected phenotypic changes. Glucose-6-phosphate (G6P) analogues 14–17 of compounds 1–4 were synthesized (Supplementary Methods and Extended Data Fig. 8) and compared for their activity; G6P-methyl-glycoside itself is inactive *in planta* and in all interactions with SnRK1 (Extended Data Fig. 8) and so its light-activated release from analogues 14–17 provided a useful control. Analogues 14–17 showed similar light-activated release parameters to compounds 1–4 (Supplementary Table 11) and relative uptake performance was similarly dependent on the identity of the light-sensitive moiety (Extended Data Fig. 8 and Supplementary Table 12–14). No toxicity was observed in any of the plants treated with up to 0.5 mM of analogues 14–17 (Supplementary Methods and Extended Data Fig. 5a–c), suggesting that the light-released moiety is benign. Critically, starch was not affected in controls treated with compound 16, the G6P analogue of T6P precursor 3.

The rate of starch synthesis in *A. thaliana* over a 12-h period (Extended Data Fig. 5f, g) indicated a flux ($0.037 \mu\text{mol min}^{-1} \text{g}^{-1}$ fresh weight) nearly three times that of water-treated controls ($0.013 \mu\text{mol min}^{-1} \text{g}^{-1}$ fresh weight). T6P is proposed to stimulate starch synthesis through redox activation²⁶ of ADP-glucose pyrophosphorylase (AGPase), a rate-limiting enzyme. While not necessarily causal, consistent with this hypothesis, plants treated with compound 3 had significantly higher AGPase activity (increased by 35%, Extended Data Fig. 5e). AGPase has been previously shown to affect starch turnover²⁷.

We also measured the number of transcripts of genes known to be associated with T6P. First, SnRK1 is a proposed target of T6P³⁰. SnRK1-induced (*TPS5*, *bZIP11* (also known as *GBF6*) and *UDPGDH* (At3g29360)) and -repressed (*TPS8* and *ASN1*) markers responded synchronously to the activation of the precursors in a manner consistent with known effects of T6P on SnRK1 activity (Extended Data Fig. 9). However, other markers (for example, *UDPGDH* (At3g29360) and *bGAL4*) showed clear temporal delay (Extended Data Fig. 9b); the observed synchronization of these 'secondary markers' only occurred after a day, which suggests that these could be later, downstream targets of T6P. Second, starch is also a proposed target of T6P^{26,31}. As starch levels were increased as a result of treatment, expression of starch biosynthetic genes was also analysed. Transcripts of *APL3*, *SS3*, *BE1* (also known as *EMB2729*) and *GBSS1* were increased up to fivefold (Extended Data Fig. 9c).

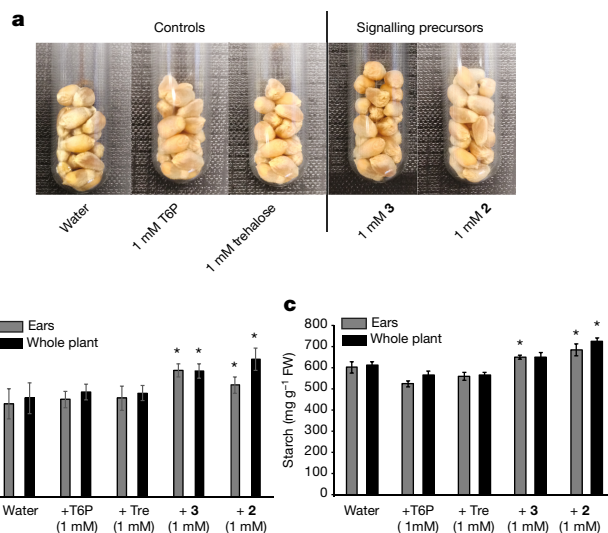


Figure 5 | Increased crop yield. **a**, Increased grain size (20 per tube) after spraying. **b**, Grain yield per plant with 1 mM of compounds 2 or 3. **c**, Starch content of the grain. * $P < 0.05$ compared to water control (Student's *t*-test). Data are shown as mean \pm s.e.m. ($n = 6$).

These data from *A. thaliana* raise the noteworthy possibility of enhanced starch synthesis in crops, potentially providing increased yield. Signalling precursors 2 and 3 were applied (0.1, 1 and 10 mM) to spring wheat (*Triticum aestivum* Cadenza), which was grown in a controlled environment representative of summer in northern Europe. Spraying occurred either to ears only or to the whole plant during the grain-filling period (5, 10, 15 and 20 days post-anthesis (the flowering period, DPA)) at mid-photoperiod. This increased grain yield per plant due to the formation of larger grain, particularly in plants treated with 1 mM of compounds 2 or 3 (Fig. 5a, b). In these grains, starch content increased 13–20% (Fig. 5c). A trend towards higher levels of starch and protein, when expressed as a percentage of component content per gram of grain, was also observed (Supplementary Table 16). Dose-response analysis showed that yield peaked at a precursor concentration of 1 mM (Extended Data Fig. 10f). Minimal spray amounts at only 10 DPA increased yield substantially at 1 and 10 mM doses (Extended Data Fig. 10g). Plants treated with compounds 2 and 3 stayed greener for longer than plants treated with water only, consistent with chlorophyll content (Extended Data Fig. 10a, b) and previous observations for genetically-enhanced T6P content²⁸. T6P release

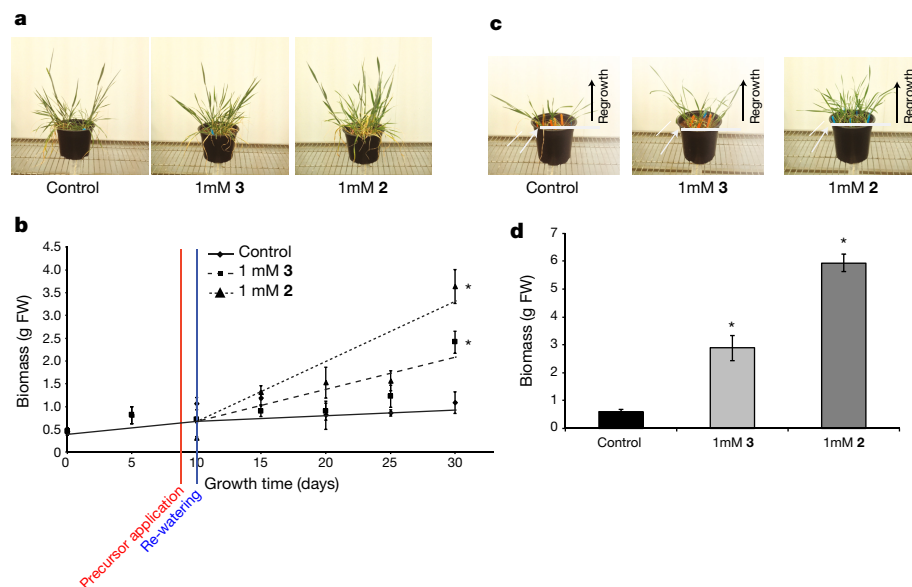


Figure 6 | Increased crop resilience. **a**, Plants after 20 d recovery following one application of 1 mM 3 or 2 1 d before rewatering. **b**, Dry weight (DW) biomass from plants in **a**. **c**, Plants after one application of 1 mM 3 or 2 one day before rewatering, cut at 5 d after rewatering, and left to regrow for 10 d. White arrow and line, cut back point. **d**, Fresh weight (FW) biomass of regrowth from **c**. **d**, * $P < 0.05$ compared to water-treated control (Student's *t*-test, $n = 6$). Data are shown as mean \pm s.e.m.

in the wheat grains treated with compounds **2** and **3** was enhanced at 5 DPA (128 nmol g⁻¹ fresh weight and 81 nmol g⁻¹ fresh weight, respectively) and further at 10 DPA (378 nmol g⁻¹ fresh weight and 300 nmol g⁻¹ fresh weight, respectively) (Extended Data Fig. 10c–e). Trehalose levels were also higher (30–70 nmol g⁻¹ fresh weight compared to endogenous levels of 13 nmol g⁻¹ fresh weight), consistent with the metabolism observed in *A. thaliana*.

Next, the effects of signalling precursors on plant resilience and recovery were analysed. Drought is still the biggest global factor limiting crop yields, even in developed countries²⁹. When 4-week-old wheat plants were sprayed with compounds **2** or **3** (30 ml, 1 mM, once) after 9 days of drought, the regrowth effects following resumption of watering 1 day after treatment were substantial (Fig. 6a, b). Regrowth of new tissue from plants cut back after drought was also higher in precursor-treated plants (Fig. 6c, d). This demonstrated both growth of new tissue (resurrection response) and salvage and growth of existing tissue (recovery response). T6P solution alone gave identical results to water (Fig. 6), consistent with the inability of T6P to enter directly into plants, further highlighting the design principles of signalling precursors.

In conclusion, we have shown here that a chemical strategy can directly control amounts of an important sugar signalling molecule *in vivo*. The collected data are consistent with the signalling action of released T6P. For example, the mass balance of added signalling precursor appears insufficient to simply act as a carbon source. That said, we do not discount other possible mechanisms behind the noteworthy traits that we have observed here. The apparent result of 'biosynthetic amplification' observed from signalling precursors is, we believe, a promising concept; we calculate here up to 50-fold 'molecular amplification' of the plant sugar product compared to the precursor. It may therefore be possible to design a self-sustaining production strategy in which a fraction of the additional starch generated by this amplification is used as a feedstock chemical for eventual synthesis of the signalling precursors themselves (Supplementary Discussion and Supplementary Table 18).

We speculate that this chemical approach also offers more temporal and strategic flexibility than genetic methods (for example, a 'pulse' to circumvent adaptation effects or in manipulating more genetically complex crops) as well as the prospect of providing an immediate boost to productivity at critical times in the plant life cycle (for example, to allow synchronicity with the sun or to rescue drought-stricken regions, Supplementary Discussion)—the potential to contribute towards global food security seems notable and immediate. Given the widespread importance of cell signalling and of carbohydrates in biology, this system may also have even wider utility.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 April 2015; accepted 28 October 2016.

Published online 14 December 2016.

- Ray, D. K., Mueller, N. D., West, P. C. & Foley, J. A. Yield trends are insufficient to double global crop production by 2050. *PLoS One* **8**, e66428 (2013).
- Lobell, D. B. & Tebaldi, C. Getting caught with our plants down: the risks of a global crop yield slowdown from climate trends in the next two decades. *Environ. Res. Lett.* **9**, 074003 (2014).
- Shewry, P. R. Wheat. *J. Exp. Bot.* **60**, 1537–1553 (2009).
- Nuccio, M. L. *et al.* Expression of trehalose-6-phosphate phosphatase in maize ears improves yield in well-watered and drought conditions. *Nat. Biotechnol.* **33**, 862–869 (2015).
- Figuerola, C. M. & Lunn, J. E. A tale of two sugars: trehalose 6-phosphate and sucrose. *Plant Physiol.* **172**, 7–27 (2016).
- Mayer, G. & Heckel, A. Biologically active molecules with a 'light switch'. *Angew. Chem. Int. Edn Engl.* **45**, 4900–4921 (2006).
- Adams, S. R. & Tsien, R. Y. Controlling cell chemistry with caged compounds. *Annu. Rev. Physiol.* **55**, 755–784 (1993).
- Givens, R. S. & Kueper, L. W. Photochemistry of phosphate-esters. *Chem. Rev.* **93**, 55–66 (1993).
- Ellis-Davies, G. C. Caged compounds: photorelease technology for control of cellular chemistry and physiology. *Nat. Methods* **4**, 619–628 (2007).

- Scheiget, J. & Roy, B. Synthesis of fluorescein phosphotriesters using photolabile protecting groups. *Synth. Commun.* **30**, 1437–1445 (2000).
- Arsilan, T., Mamaev, S. V., Mamaeva, N. V. & Hecht, S. M. Structurally modified firefly luciferase, effects of amino acid substitution at position 286. *J. Am. Chem. Soc.* **119**, 10877–10887 (1997).
- Rønnow, T. E., Meldal, M. & Bock, K. Gram-scale synthesis of alpha, alpha-trehalose 6-monophosphate and alpha, alpha-trehalose 6,6'-diphosphate. *Carbohydr. Res.* **260**, 323–328 (1994).
- Meldal, M., Christensen, M. K. & Bock, K. Large-scale synthesis of D-mannose 6-phosphate and other hexose 6-phosphates. *Carbohydr. Res.* **235**, 115–127 (1992).
- Yang, M., Brazier, M., Edwards, R. & Davis, B. G. High-throughput mass-spectrometry monitoring for multisubstrate enzymes: determining the kinetic parameters and catalytic activities of glycosyltransferases. *ChemBioChem* **6**, 346–357 (2005).
- Clarke, E. D. & Delaney, J. S. Physical and molecular properties of agrochemicals: an analysis of screen inputs, hits, leads, and products. *Chimia* **57**, 731–734 (2003).
- Kirkwood, R. C. *Target Sites for Herbicide Action*. (Springer, 1991).
- Delatte, T. L. *et al.* Determination of trehalose-6-phosphate in *Arabidopsis* seedlings by successive extractions followed by anion exchange chromatography-mass spectrometry. *Anal. Biochem.* **389**, 12–17 (2009).
- Bjarnholt, N., Li, B., D'Alvise, J. & Janfelt, C. Mass spectrometry imaging of plant metabolites—principles and possibilities. *Nat. Prod. Rep.* **31**, 818–837 (2014).
- Cha, S. *et al.* Direct profiling and imaging of plant metabolites in intact tissues by using colloidal graphite-assisted laser desorption/ionization mass spectrometry. *Plant J.* **55**, 348–360 (2008).
- Cheng, J. & Winograd, N. Depth profiling of peptide films with TOF-SIMS and a C60 probe. *Anal. Chem.* **77**, 3651–3659 (2005).
- Nunes, C. *et al.* Regulation of growth by the trehalose pathway: relationship to temperature and sucrose. *Plant Signal. Behav.* **8**, e26626 (2013).
- Szeczowska, M. *et al.* Metabolic fluxes in an illuminated *Arabidopsis* rosette. *Plant Cell* **25**, 694–714 (2013).
- Schluepmann, H. *et al.* Trehalose mediated growth inhibition of *Arabidopsis* seedlings is due to trehalose-6-phosphate accumulation. *Plant Physiol.* **135**, 879–890 (2004).
- Smith, A. M. & Zeeman, S. C. Quantification of starch in plant tissues. *Nat. Protocols* **1**, 1342–1345 (2006).
- Givens, R. S., Weber, J. F., Jung, A. H. & Park, C. H. New photoprotecting groups: desyl and p-hydroxyphenacyl phosphate and carboxylate esters. *Methods Enzymol.* **291**, 1–29 (1998).
- Kolbe, A. *et al.* Trehalose 6-phosphate regulates starch synthesis via posttranslational redox activation of ADP-glucose pyrophosphorylase. *Proc. Natl Acad. Sci. USA* **102**, 11118–11123 (2005).
- Hädrich, N. *et al.* Mutagenesis of cysteine 81 prevents dimerization of the APS1 subunit of ADP-glucose pyrophosphorylase and alters diurnal starch turnover in *Arabidopsis thaliana* leaves. *Plant J.* **70**, 231–242 (2012).
- Pellny, T. K. *et al.* Genetic modification of photosynthesis with *E. coli* genes for trehalose synthesis. *Plant Biotechnol. J.* **2**, 71–82 (2004).
- Boyer, J. S. Plant productivity and environment. *Science* **218**, 443–448 (1982).
- Zhang, Y. *et al.* Inhibition of SNF1-related protein kinase1 activity and regulation of metabolic pathways by trehalose-6-phosphate. *Plant Physiol.* **149**, 1860–1871 (2009).
- Martins, M. C. M. *et al.* Feedback inhibition of starch degradation in *Arabidopsis* leaves mediated by trehalose 6-phosphate. *Plant Physiol.* **163**, 1142–1163 (2013).
- Billington, D. C. Recent developments in the synthesis of myo-inositol phosphates. *Chem. Soc. Rev.* **18**, 83–122 (1989).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the BBSRC Selective Chemical Intervention in Biological Systems initiative (grant reference BB/D006112/1), the BBSRC Sparking Impact initiative and ICL Innovations for funding. We thank R. H. Bromilow, S. Powers, E. Tobolkina, for advice and J. Wickens for technical assistance. NiCE-MSI is supported by the 3D NanoSIMS and AIMS-HIGHER projects of the Chemical and Biological programme of the National Measurement System of the UK Department of Business, Innovation and Skills. B.G.D. was a Royal Society Wolfson Research Merit Award recipient during the period of research. Rothamsted Research receives strategic funding from the BBSRC.

Author Contributions C.A.G., R.S. and Y.G. are joint first authors. C.A.G., R.S., Y.G. and L.F.P. performed experiments. R.S., Y.G. and M.K.P. synthesized compounds. Y.G. and M.K.P., I.S.G., R.T.S., J.B. and B.G.D. performed and/or analysed the mass spectrometry imaging. Y.G. and B.G.D. performed and/or analysed the tandem mass spectrometry. C.A.G., R.S., Y.G., L.F.P., M.J.P. and B.G.D. designed and analysed the experiments. M.J.P. and B.G.D. wrote the paper. All authors read and commented on the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.J.P. (matthew.paul@rothamsted.ac.uk) and B.G.D. (Ben.Davis@chem.ox.ac.uk).

METHODS

Synthesis of signalling-precursor compounds 1–4. 1*H*-tetrazole solution (0.45 M in CH₃CN) (0.6 ml, 0.24 mmol, 2.0 equiv.) was added into a stirred solution of compound **12** (100 mg, 0.12 mmol, 1 equiv.) and *bis*-(2-nitrobenzyl)-*N*, *N*-diisopropylphosphoramidite (compound **9**; 78.3 mg, 0.18 mmol, 1.5 equiv.) in anhydrous CH₂Cl₂ (5 ml) under an argon atmosphere at 0 °C. The resulting reaction mixture was stirred at 0–5 °C and progress of the reaction was monitored by thin-layer chromatography (petroleum ether:ethyl acetate, 8:2) and mass spectrometry. After complete disappearance of starting material (1 h), tBuOOH (0.1 ml) was added at 0 °C and stirring was continued for another 30 min. After 30 min the reaction mixture was concentrated *in vacuo* and the residue was suspended in methanol (2 ml) and stirred in the presence of 30 mg of Dowex-H⁺ resin for 1 h at room temperature to globally remove trimethylsilyl groups. Dowex-H⁺ was removed through filtration and the filtrate was concentrated, which on flash chromatography (water:isopropanol:ethyl acetate, 1:2:8) purification yielded compound **1** (70 mg) in 87% isolable yield. Similar reaction protocols were used for the synthesis of compounds **2** and **3**. Compound **4** was obtained when a stirred solution of **12** (100 mg, 0.12 mmol) in pyridine (2 ml) at room temperature was treated with POCl₃ (0.012 ml, 0.132 mmol) for 10 min followed by addition of 4,5-dimethoxy-2-nitrobenzyl alcohol (76.7 mg, 0.36 mmol) and continuous stirring for 1 h. The resulting reaction mixture was concentrated *in vacuo* to yield a crude product mixture, which was treated with Dowex-H⁺ (30 mg) in methanol (2 ml). After filtration, concentration *in vacuo* and flash chromatography purification yielded compound **4** (45 mg, 62%) as a pure sticky solid. For additional details see Supplementary Methods.

***In planta* uptake of signalling-precursor compounds and release of trehalose-6-phosphate and metabolites.** *In planta* uptake was carried out using *A. thaliana* plantlets. *A. thaliana* (Columbia 0) seeds were surface-sterilized for 10 min in 10% sodium hypochlorite, 0.01% Triton X-100 and then copiously washed with sterile water and stratified for 3 d at 4 °C. Seeds were sown onto 0.5 ml solid medium (0.5 × Murashige and Skoog medium with Gamborg's vitamins (Sigma P0404), 0.5% sucrose and 0.5% agar) in 0.5 ml Eppendorf tubes, pierced in the bottom with a tiny hole. The tubes were arrayed in hand-cut polystyrene racks in Phyta trays (Sigma) and floated on liquid medium (same as solid medium but lacking sucrose and agar). Plantlets were grown under the following conditions: 12 h day under Philips master TL-D 840/58W fluorescent lights outputting 250 μmol m⁻² s⁻¹, with 23 °C day and 18 °C night temperatures. At 18 days after sowing the liquid medium was removed and the tubes were sealed with electrician's tape. All plants were topped up with 0.5 × Murashige and Skoog medium with no sucrose.

Plants were treated with compounds by adding 10 μl of a 50 mM stock prepared in water or 1% DMSO to the agar medium, avoiding contact with aerial parts. The final concentration of the precursor compound in the agar medium was 1 mM. After a certain period of time (after 24 h, 48 h and 72 h) the aerial part was harvested carefully, weighed and extracted in H₂O:MeOH (1:1) under liquid nitrogen. The crude fresh plant extract thus obtained was analysed by mass spectrometry and HPLC.

For *in planta* T6P release experiments, compound-treated plants were exposed to UV-light treatment after 72 h. UV treatments consisted of: (a) 8-h exposure to natural daylight; (b) 8-h exposure to a 100 W UV spotlight (BlackRay B-100AP) at a distance of 18 cm; (c) 8-h exposure to an 8 W UV bulb (365 nm, Gelman transilluminator Model 51438) at a distance of 6 cm; or (d) exposure for two 8-h periods to 8 W UV treatments were in addition to normal growth lights. Control plants (except for daylight treatment) were treated under the same conditions but without UV light. At the end of the day of exposure to UV or visible light, the aerial parts of the plants were quickly harvested, weighed and frozen in liquid nitrogen. For starch extractions, a moderate light regime was selected of a single 8 h exposure to 8 W UV light. After irradiation plants were returned to the growth room for a further day (day 5) to recover after light treatment and to respond to altered T6P levels before being harvested as above. Frozen tissue was stored at –80 °C until extracted.

Harvested plant material was extracted by liquid/liquid extraction (LLE) followed by solid phase extraction (SPE) for T6P analysis¹⁷. For LLE/SPE extractions around 25 mg plant tissue was used, pooled from several plants. Samples were reconstituted in 50 μl of H₂O:MeOH (1:1) and 10 μl was used for T6P analysis; T6P release was determined using quantitative HPLC–MS (Quattro, Waters), with 2-deoxy-glucose-6-phosphate as a calibration internal standard.

Liquid chromatography tandem mass spectrometry (LC–MS/MS) was used to confirm the identity of disaccharide monophosphates via fragmentation pattern analysis performed on a Waters Xevo G2-S QToF (quadrupole time-of-flight) mass spectrometer coupled to a Waters Acquity Ultra Performance Liquid Chromatography (UPLC) system, and a Waters Micromass Quattro micro API Mass Spectrometer coupled to a Waters 1525 μ binary HPLC pump and a Waters 2777 auto sampler using a SIELC Primesep SB column. Solvent A (0.1% formic

acid in H₂O) and solvent B (1.0% formic acid in H₂O:CH₃CN (75:25)), were used as the mobile phase at a flow rate of 0.4 ml min⁻¹. For the Xevo G2-S QToF MS, the electrospray source was operated with a capillary voltage of 2.0 kV and a cone voltage of 30 V. Nitrogen was used as the desolvation gas at a total flow of 800 l h⁻¹. The intact molecular ion of T6P was detected as *m/z* 421.0759 (C₁₂H₂₂O₁₄P, calculated as 421.0753) in a negative ion mode. The time of flight (ToF) tandem mass spectrum of the parent ion 421.00 was then obtained in a negative ion mode for the *m/z* range from 50 to 500 using optimized collision energy of 20 eV. For the Quattro micro API MS, the electrospray source was operated with a capillary voltage of 3.0 kV and a cone voltage of 40 V. The quadrupole tandem mass spectra of the parent ion 421.0 of T6P and S6P were obtained in a negative ion-mode for the *m/z* range from 50 to 500 using collision energy of 20 eV. With the reference to standard fragmentation patterns, tracking of the fragment ions of T6P in the plant sample was also performed by quadrupole tandem mass spectra. The five most intense *m/z* peaks recorded in the MS/MS spectrum of T6P; *m/z* 78.3, 96.4, 138.6, 240.9 and 421.0 (unfragmented) were also selected for multiple reaction monitoring (MRM) and cross-referenced with selected ion recordings (SIR) for the intact molecular ion *m/z* 421.0.

For seedling liquid culture, seeds of *A. thaliana* were grown in liquid culture as described previously²¹. Once the seedlings were 7 days old, oNPE-T6P (compound **3**) or DMNB-T6P (compound **2**) were added to the growth medium to a final concentration of 1 mM. Plants were left under growth lights to take up the compounds for 24 h. To facilitate precursor release, plants were placed under 23 μmol m⁻² s⁻¹ UV for 2 h, after which they were returned to previous environmental conditions. Samples were taken for analysis before addition of the compound, 1 day after addition, after 30, 60 and 120 min during UV treatment, and sampled again at 1 and 2 d after UV treatment. Samples were weighed, snap-frozen and stored at –80 °C.

For enzymatic sugar analysis, sugars were extracted from 5–10 mg of *A. thaliana* ground under liquid nitrogen, 1 ml of 80% was added and the sample was heated at 100 °C for 1 h, samples were centrifuged for 10 min at 13,000g to remove debris. The samples were added to assay buffer²⁴. Enzymatic reactions were performed as described previously³² using hexokinase, glucose-6-phosphate dehydrogenase, phosphoglucose isomerase and invertase from Sigma-Aldrich (H4502, G8404, P5381 and I9274, respectively). Two technical replicates were carried out for each sample, a total of three biological replicates were analysed. See Supplementary Methods for further details.

Extraction and measurement of starch *in planta*. Three or four chemical- and UV-light-treated plantlets were pooled and weighed (fresh weight, 70–100 mg) for each biological replicate. Extraction was based on literature methods²⁴. Samples were ground in liquid nitrogen to a fine powder in a mortar. The powder was rapidly extracted with 1 ml 80% ethanol at 80 °C, followed by 2 × 0.5 ml to rinse, samples were then transferred to a 2-ml eppendorf at 100 °C and heated for 2–3 min until just boiling. Tubes were transferred to a water bath at 80 °C while other samples were accumulated. Samples were centrifuged at 13,000g for 10 min to collect all solid material. The pellet was extracted twice more with 2 ml, 80% hot ethanol. The pellet was washed with 1 ml water, the supernatant removed and 100 μl water added. The pellet was homogenized to a smooth consistency with an Eppendorf micropestle before being made up to final volume of 500 μl with water. Samples were heated at 100 °C for 10 min to gelatinize starch granules. Duplicate aliquots (100 μl) were removed and digested with α-amylase (2 U) and amyloglucosidase (6 U) in 0.05 M sodium acetate pH 4.8 for 4 h at 37 °C. Control digests without enzyme were also set up. Glucose released from digested starch was measured using an enzymatic assay coupled to the reduction of NADP to NADPH³³ and adapted for a microtitre plate reader. 10–20 μl of digest was assayed in triplicate. Starch content is expressed as hexose equivalents per gram fresh weight. See Supplementary Methods for further details.

SnRK1 activity. Kinase activities were determined by measuring the incorporation of radiolabelled phosphate into the 'AMARA' peptide (AMARAASAAALARRR; Biomol International) substrate and were carried out as described previously³⁰.

ADP-glucose pyrophosphorylase activity. Enzyme activity was measured as described previously³⁴.

Application of precursors to wheat. Spring wheat (*T. aestivum* Cadenza) seeds were sown in Rothamsted standard compost mix and grown in controlled environment conditions with a photoperiod of 16 h light, 8 h dark, day/night temperatures of 20 °C/16 °C, photon-flux density of 600 μmol m⁻² s⁻¹, and ambient relative humidity. Once the plants had reached anthesis, solutions of oNPE-T6P (compound **3**) and DMNB-T6P (compound **2**) (0.1, 1 or 10 mM) as well as control solutions, water, 1 mM T6P or 1 mM trehalose were made up in distilled water with 0.1% TWEEN-20. At 5, 10, 15 and 20 days after anthesis, either the ears, or the whole plant were sprayed individually with 5 or 50 ml of precursor, respectively. Leaf samples were taken at 5, 10, 15, 20 and 25 days after anthesis for chlorophyll

content analysis, grain was harvested at maturity for analysis. Chlorophyll content of leaves was measured by methanol extraction and spectrophotometry³⁵. Starch content of grain was measured enzymatically²⁴ and protein content was measured by Bradford's assay³⁶.

For the drought treatment, vegetative Cadenza wheat plants were grown in the same compost and environments as above. Once the plants had reached Feekes stage 4, water was withheld for 10 days. On day 9, 30-ml, 1 mM solutions of oNPE-T6P (compound 3) and DMNB-T6P (compound 2) were applied to all above-ground biomass, on day 10 the watering schedule was reinstated. Plants were harvested to measure biomass production every 5 days for 30 days after rewatering. Both experiments were completed in biological replicates of six.

For quantification of T6P, trehalose and sucrose in wheat samples, the harvested wheat grains were weighed, snap-frozen and stored at -80°C . Wheat grain was ground to a fine powder in liquid nitrogen and the sugars were extracted by LLE for T6P, trehalose and sucrose analysis using the same LC-MS quantification method as for *A. thaliana*.

For minimal spray application, spring wheat (*T. aestivum* Cadenza) seeds were sown in Rothamsted standard compost mix and grown in controlled environment conditions with a photoperiod of 16 h light, 8 h dark, day/night temperatures of $20^{\circ}\text{C}/16^{\circ}\text{C}$, photon-flux density of $600\mu\text{mol m}^{-2}\text{s}^{-1}$ and ambient relative humidity. Once the plants had reached anthesis, solutions of compounds 2 or 3 (1 mM and 10 mM) and a water control, were made up in distilled water with 0.1% TWEEN-20. At 10 days after anthesis, the top 20 cm of above ground biomass encompassing ears and flag leaves were sprayed individually with 25 ml of the compounds. Grain from individual ears was harvested at maturity for analysis. All wheat experiments were repeated twice, with 3 technical and 6 biological replicates completed at each stage of analysis.

RNA extraction, cDNA synthesis and qRT-PCR. Total RNA was extracted from 50 mg snap-frozen leaf tissue from *A. thaliana* Columbia using the Ribopure Kit (Ambion) according to the manufacturer's instructions. RNA was quantified using a Nanodrop spectrophotometer and integrity of RNA was visualized using denaturing agarose gel electrophoresis³⁷. DNA was removed using RQ1 RNase-free DNase (Promega). cDNA was synthesized using SuperScript III First-Strand Synthesis System (ThermoFisher Scientific) using $2\mu\text{g}$ of total RNA and oligo-dT primers according to the manufacturer's instructions. Gene expression was quantified using SYBR Green chemistry on a Real-Time PCR system 7500 (Applied Biosystems). Total reaction size was $20\mu\text{l}$, containing $10\mu\text{l}$ SYBR Green Jumpstart Taq ReadyMix (Sigma Aldrich), $2\mu\text{l}$ cDNA and 0.5 mM primers. PCR used an initial denaturation stage of 95°C for 2 min, followed by 40 cycles of 95°C for 15 s, 60°C for 1 min. The specificity of products was confirmed by performing a temperature gradient analysis of products at temperatures ranging from 55°C to 95°C with 0.5°C increments. Two technical replicates were completed for each sample, a total of three biological replicates were analysed. Relative quantification of gene expression was performed using the Livak method using ubiquitin-transferase family protein as the reference gene. Primers used for SnRK1 marker gene expression, and starch gene expression are listed in Supplementary Tables 19 and 20, respectively.

Mass spectrometry imaging methods. *A. thaliana* used in ToF-SIMS (time-of-flight secondary ion mass spectrometry) imaging were grown in Petri dishes on $0.5\times$ Murashige and Skoog medium with 0.8% agar for 10 d, with a photoperiod of 16 h light, 8 h dark, day/night temperatures of $23^{\circ}\text{C}/18^{\circ}\text{C}$ and photon-flux density of $250\mu\text{mol m}^{-2}\text{s}^{-1}$. Plants were then transferred to Petri dishes containing the same media supplemented with 1 mM of either compound 2 or 3 for 24 h, during

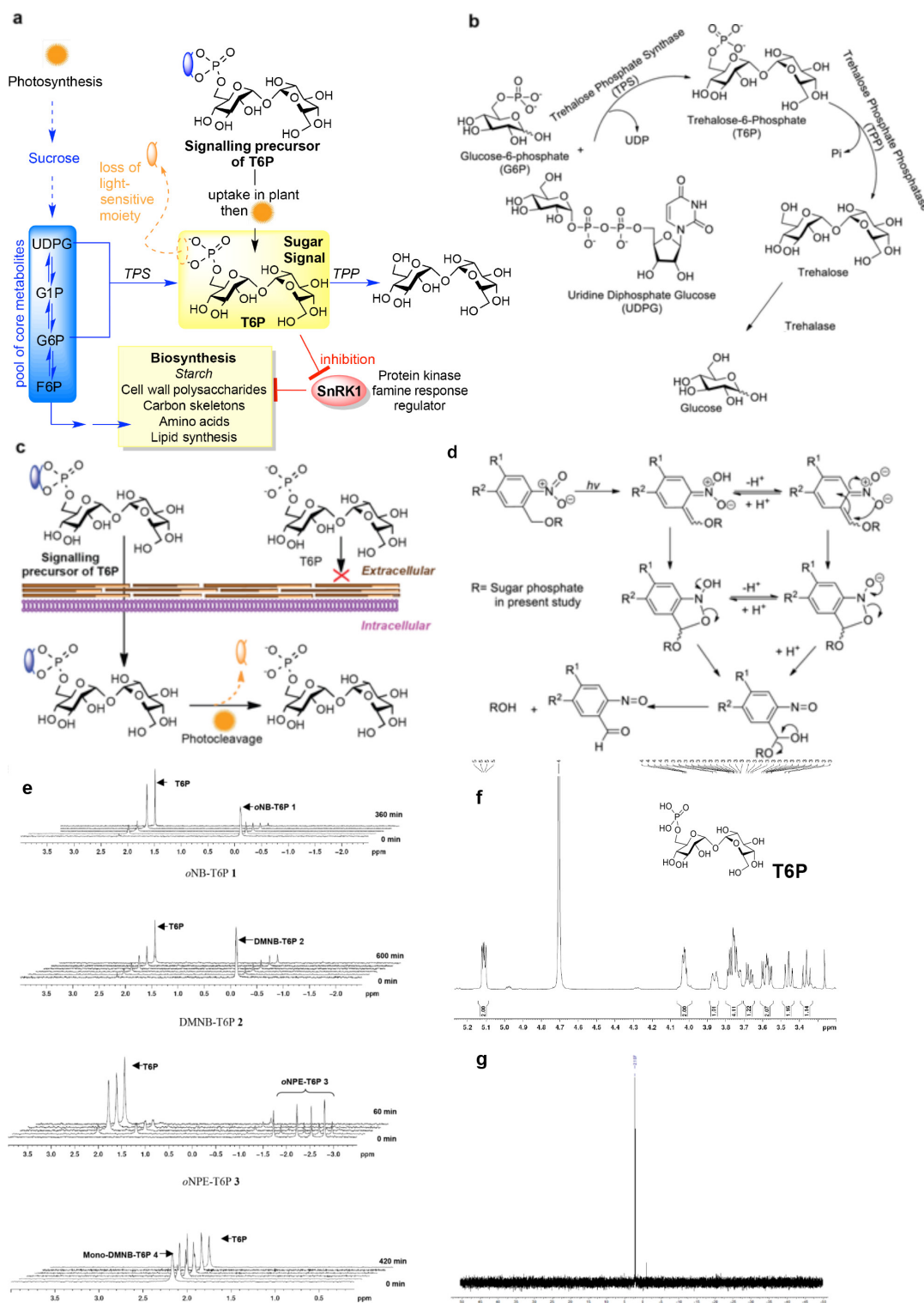
which they remained under the previously stated growth conditions. After 24 h, the plants were exposed to UV light at $23\mu\text{mol m}^{-2}\text{s}^{-1}$ for 2 h to facilitate T6P release. Plants were left for 2 h to recover, frozen and dehydrated in a vacuum chamber before mass spectrometry imaging analysis. Reference materials were drop-dried on clean substrates. The ToF-SIMS mass spectrometry imaging analysis was performed with the ToF-SIMS IV mass spectrometer (IONTOF, Muenster, Germany) from three leaves, a control, and one leaf each treated with compounds 2 or 3. A pulsed 25-keV Bi_3^{+} primary ion source was used as the analysis beam (pulse width, 23 ns, mass resolution, $(m/\Delta m)$, 5,000). Mass spectra of the reference material were obtained in positive and negative ion mode at a primary ion dose of 1.1×10^{11} ions cm^{-2} . The leaf ion images were also collected in both polarities with a dose of 5.4×10^{10} ions cm^{-2} . An electron flood gun was employed for charge compensation during the data acquisition. Mass spectrometry data was analysed in ION-TOF SurfaceLab 6.4 software and further processed in MATLAB and Origin Pro. Known melissic acid markers¹⁹ (m/z 435.4, $\text{C}_{30}\text{H}_{59}\text{O}$, $[\text{M}-\text{OH}]^{-}$ and 451.4 $\text{C}_{30}\text{H}_{59}\text{O}_2$, $[\text{M}-\text{H}]^{-}$) were used. See Supplementary Methods for further details.

MALDIMS imaging data were acquired using identically prepared leaf samples with a modified QSTAR XL Qq-ToF instrument (Sciex, Ontario, Canada) fitted with a Nd:YAG laser (Elforlight Ltd, Daventry, UK) operated at 1,000 kHz in positive ion mode with a fluence of approximately 205 J m^{-2} and a pixel size of $200\mu\text{m}\times 200\mu\text{m}$. The QSTAR was operated in continuous raster sampling mode. The sample was affixed to a stainless steel target plate with double-sided tape and sprayed with CHCA (5 mg ml^{-1} CHCA in 80% methanol, 0.1% TFA) using automated spray deposition (TM Sprayer, HTX Technologies, Carrboro, USA). Data were converted from the proprietary .wiff format into .mzML using AB MS Data Converter version 1.3 (Sciex). These mzML files were then converted to .imzML using imzMLConverter³⁸ and processed in custom-made MATLAB software (version R2014b, Math Works Inc, USA). Images are created by summing across the full-width, half-maximum of the peak of interest to give the intensity within each corresponding pixel. See Supplementary Methods for further details.

Statistical methods. ANOVA was applied to data to test for differences between treatments. A natural log transformation was used where necessary to ensure constant variance. The GENSTAT statistical system was used for this analysis.

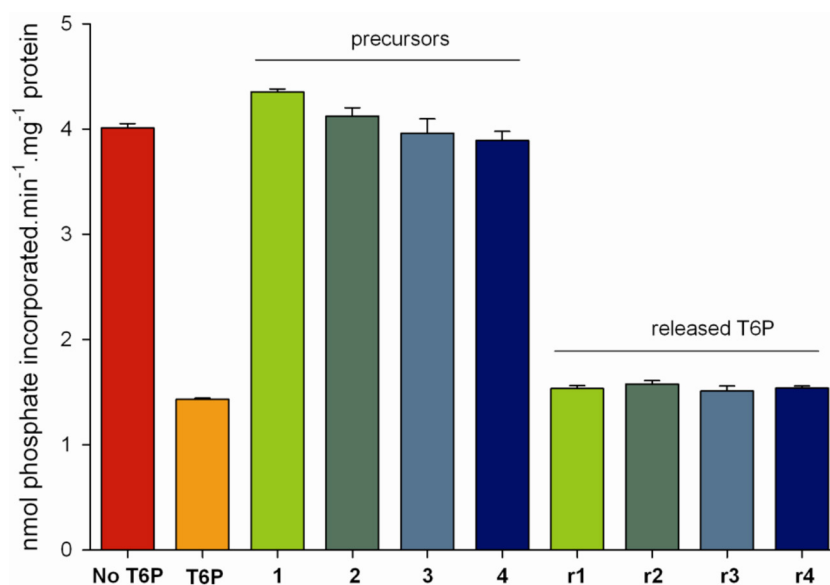
Data availability. Primary data for Figs 2, 4, 5 and 6 are provided as spreadsheets. All other data are available on request.

32. Spackman, V. M. T. & Cobb, A. H. An enzyme-based method for the rapid determination of sucrose, glucose and fructose in sugar beet roots and the effects of impact damage and postharvest storage in clamps. *J. Sci. Food Agric.* **82**, 80–86 (2002).
33. Jones, M. G. K., Outlaw, W. H. & Lowry, O. H. Enzymic assay of 10 to 10 moles of sucrose in plant tissues. *Plant Physiol.* **60**, 379–383 (1977).
34. McKibbin, R. S. *et al.* Production of high-starch, low-glucose potatoes through over-expression of the metabolic regulator SnRK1. *Plant Biotechnol. J.* **4**, 409–418 (2006).
35. Ritchie, R. J. Consistent sets of spectrophotometric chlorophyll equations for acetone, methanol and ethanol solvents. *Photosynth. Res.* **89**, 27–41 (2006).
36. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
37. Sambrook, J. & Russell, D. *Molecular Cloning: A Laboratory Manual*, 3rd Edn (Cold Spring Harbor Laboratory Press, 2000).
38. Race, A. M., Styles, I. B. & Bunch, J. Inclusive sharing of mass spectrometry imaging data requires a converter for all. *J. Proteomics* **75**, 5111–5112 (2012).



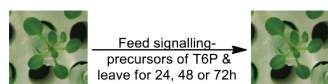
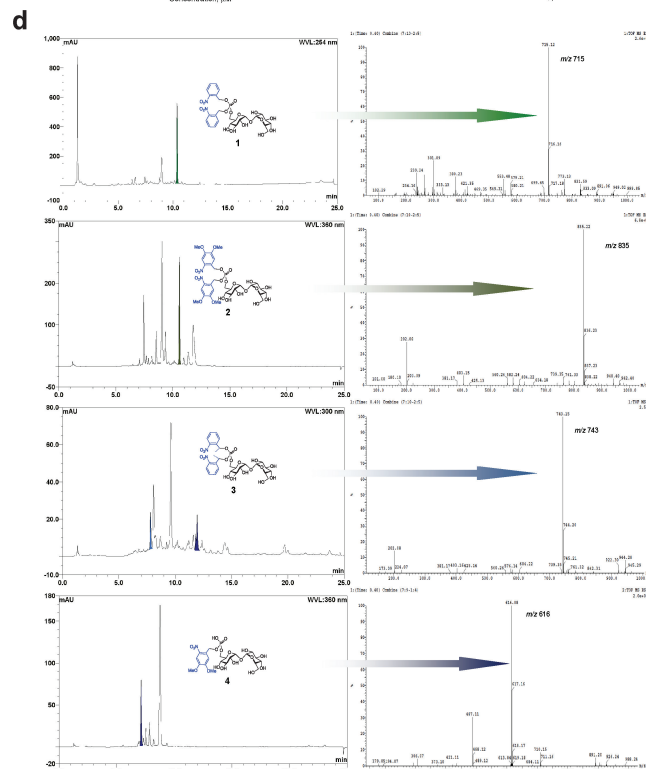
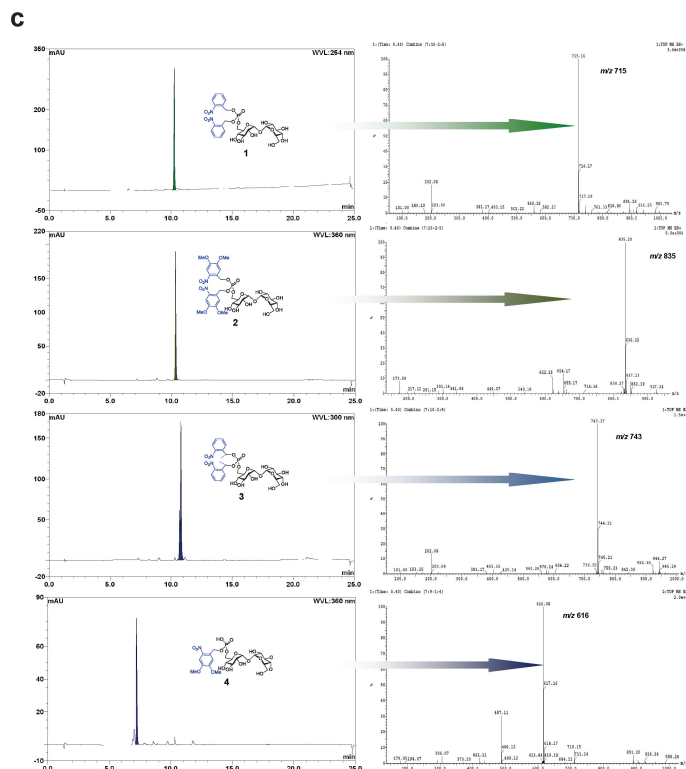
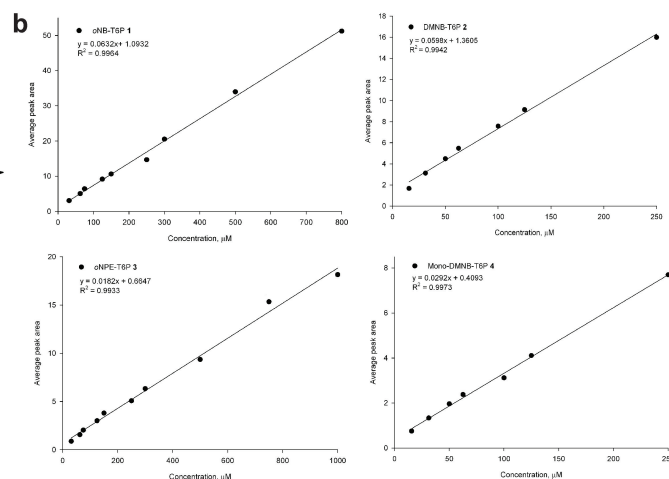
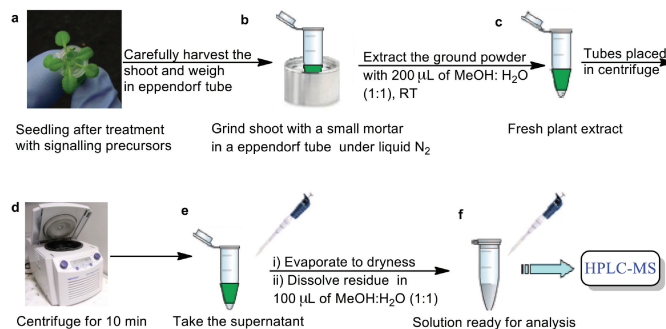
Extended Data Figure 1 | The central role of T6P in plants and design of a chemical strategy for the control of T6P production. **a**, Photosynthesis generates sucrose, which is translocated to growing regions of the plant. Inside the cell, a pool of core metabolites are substrates for biosynthetic processes that determine growth and productivity. T6P is synthesized from UDPG and G6P by trehalose 6-phosphate synthase (TPS) and therefore reflects the abundance of sucrose. It is broken down by trehalose phosphate phosphatase (TPP). Increasing T6P stimulates starch synthesis and inhibits SnRK1, a protein kinase central to energy conservation and survival during energy deprivation. Inhibition of SnRK1 by T6P thus diverts carbon skeleton consumption into biosynthetic processes. **b**, The trehalose biosynthetic pathway. **c**, T6P is plant-impermeable.

Plant-permeable variants allowed subsequent photo-activated release. **d**, Generalized mechanism of light-activated release of precursors. **e**, Release of T6P by light irradiation from signalling precursors 1–4 *in vitro*. ^{31}P nuclear magnetic resonance spectroscopy at different time points of light irradiation confirming the activation of signalling precursors (1–4) and release of T6P. Time points for compound 1: 0, 30, 60, 150 and 360 min; for compound 2: 0, 60, 120, 300, 420 and 600 min; for compound 3: 0, 15, 30, 45 and 60 min; and for compound 4: 0, 60, 120, 240, 360 and 420 min. **f, g**, ^1H (f) and ^{31}P (g) nuclear magnetic resonance spectra, after complete photolysis of the signalling precursor confirming the release of T6P.



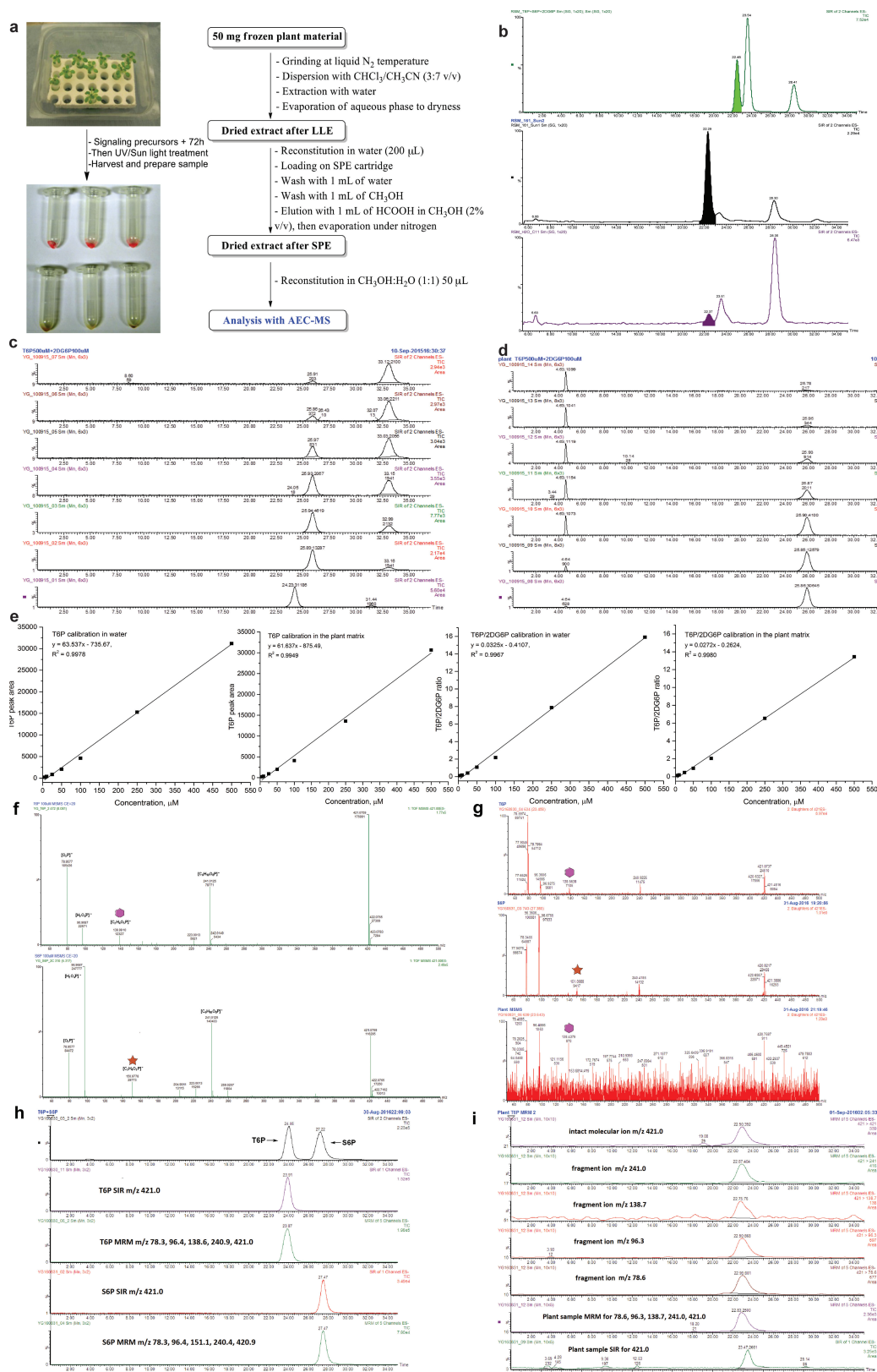
Extended Data Figure 2 | Inhibition of SnRK1. Signalling precursors (1–4), T6P released from 1–4 (r1, r2, r3, r4) and T6P standard (T6P) were tested for inhibition of SnRK1 activity. T6P (0.26 mM) inhibits SnRK1 activity to approximately 36% of the original activity. Signalling-precursor compounds show no such inhibition, whereas UV-released compounds show the same inhibition as free T6P. SnRK1 activity was

determined by the level of incorporation of phosphate into a peptide substrate ($\text{min}^{-1} \text{mg}^{-1} \text{protein}$). (Data are shown as mean \pm s.e.m.; $n = 3$). The activities of assays treated with precursors or released T6P were not significantly different from their controls ($P < 0.001$, LSD) using one-way ANOVA of data transformed with a natural log scale.

a Treatment**Analysis**

Extended Data Figure 3 | *In planta* uptake analysis of signalling precursors 1–4. **a**, Schematic of protocol used for uptake analysis. **b**, Calibration curves for oNB-T6P (1), DMNB-T6P (2), oNPE-T6P (3) and mono-DMNB-T6P (4), respectively. Data are shown as mean \pm s.e.m.; ($n = 2$). **c**, HPLC (left) and mass spectrometry (right) data, $[M + Na]^+$

or $[M - H]^-$, of pure signalling precursors 1–4. **d**, HPLC (left) and mass spectrometry (right) data, $[M + Na]^+$ or $[M - H]^-$, of plant samples after treatment with signalling precursors 1–4. In 3, the partially uncaged molecule also accumulated and was detected (coloured light blue).



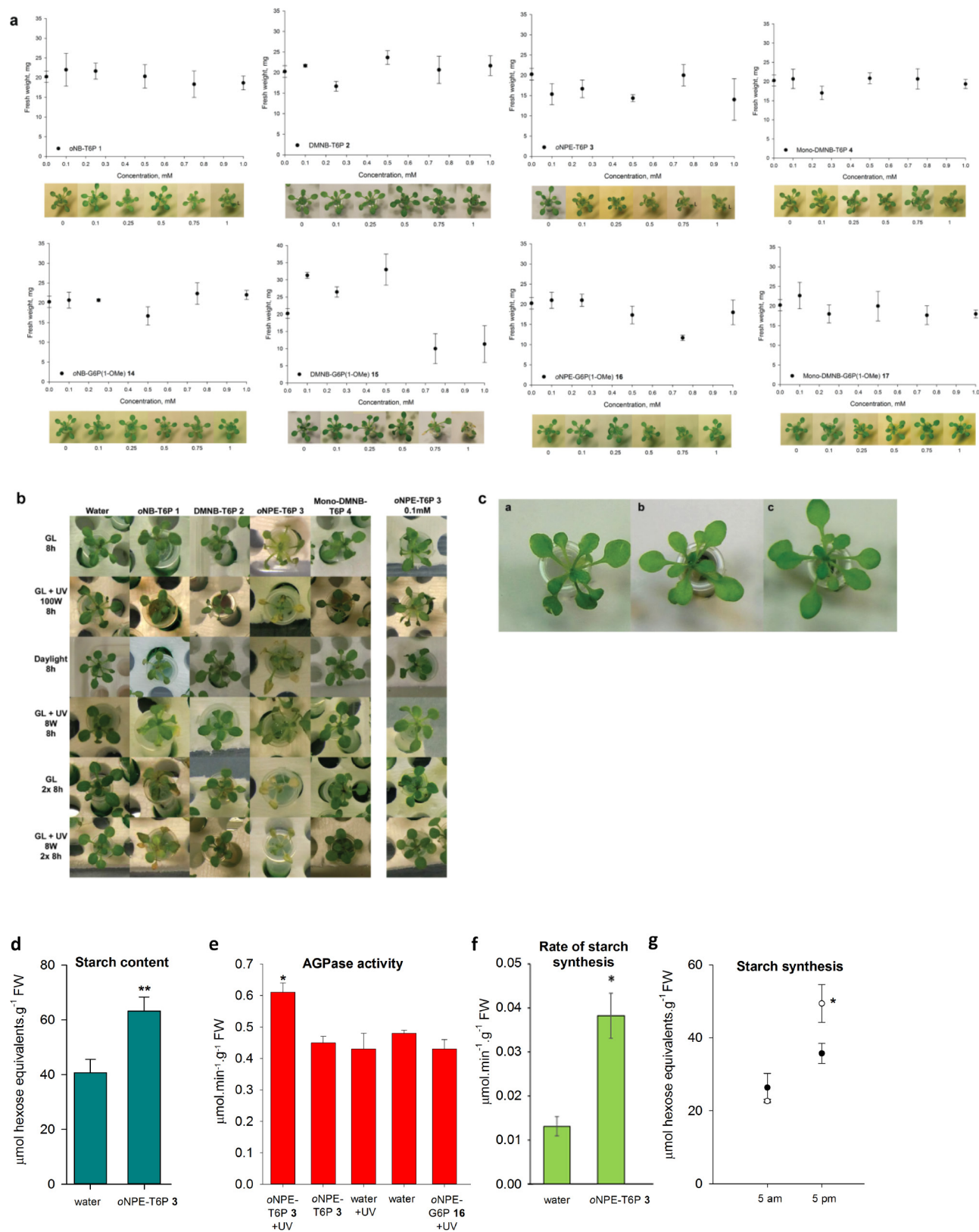
Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | Extraction and quantification of T6P.

a, Schematic of the protocol used for preparation of samples for T6P quantification. LLE, liquid/liquid extraction; SPE, solid phase extraction; AEC-MS, anion exchange chromatography-mass spectrometry.

b, Liquid chromatograms of T6P, S6P and 2DG6P separation (top) using conditions optimized in Supplementary Table 8 (entry 7) and the representative LC-MS chromatograms of extraction samples treated with signalling precursors (middle) and water control (bottom). **c, d**, Liquid chromatograms of different concentrations of T6P (500, 250, 100, 50, 25, 10 or 5 μM) using a constant concentration (100 μM) of 2DG6P as an internal standard. **e**, Resulting calibration curves of the T6P peak area and T6P/2DG6P ratio against T6P concentrations (in μM) in water as well as in the plant matrix. **f-h**, LC-MS/MS analysis of T6P and S6P from samples of plant treated with compound **2**. **f**, Fragmentation patterns of T6P (top)

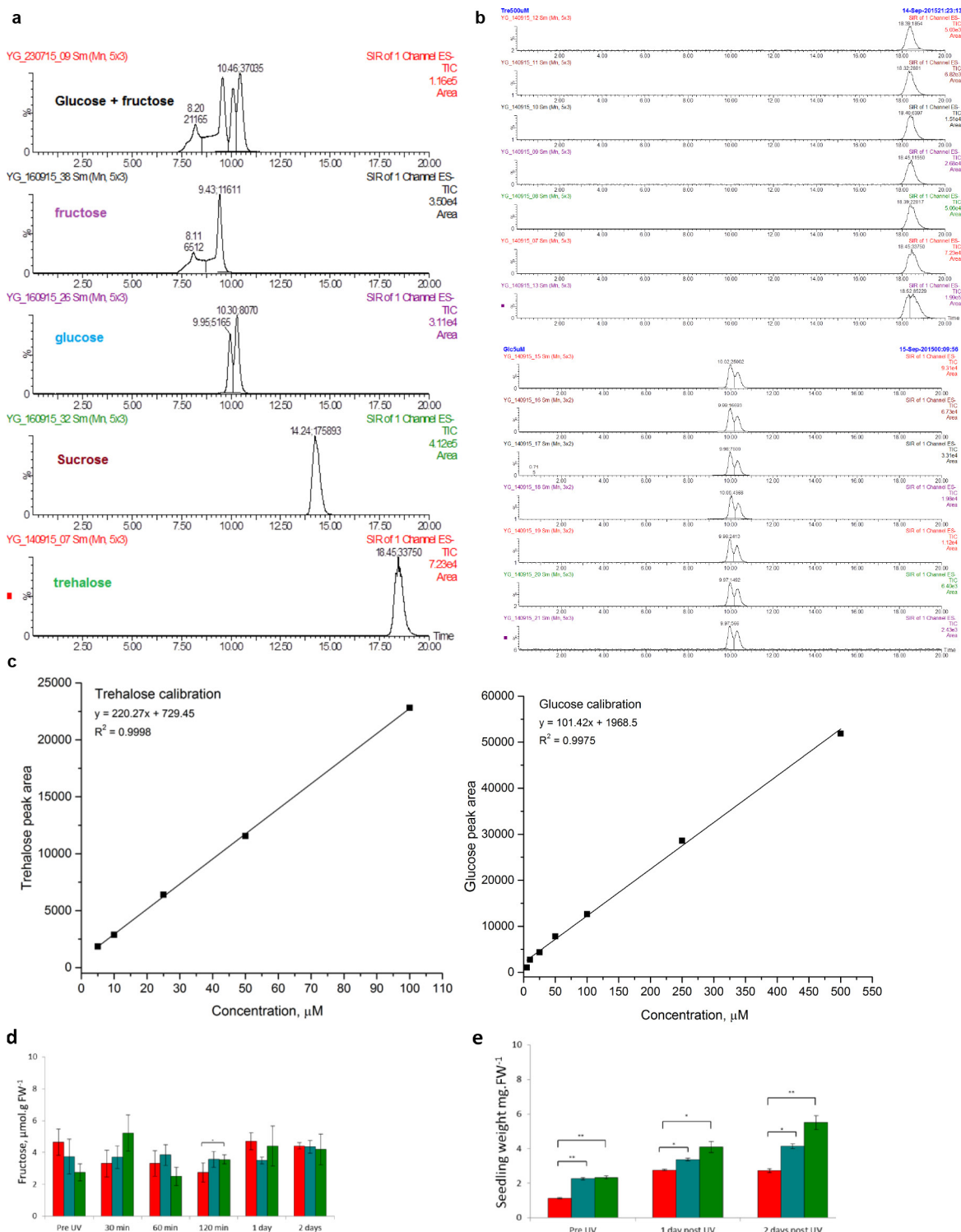
and S6P (bottom) by quadrupole time-of-flight tandem mass spectrometry (QToF-MS/MS) in negative ion mode. **g**, Fragmentation patterns of T6P (top) and S6P (middle) by triple quadrupole tandem mass spectrometry (QQQ-MS/MS) in negative ion mode and the T6P fragment ions tracking in the plant matrix (bottom). **h**, HPLC chromatograms of T6P/S6P by selected ion recording (SIR) of the intact molecular ion (m/z 421.0) and multiple reaction monitoring (MRM) of the fragment ions give the same retention time for each compound. **i**, The LC-MS quantification method through SIR and LC-MS/MS quantification method through MRM of the T6P level in the DMNB-T6P **2**-treated plant sample. From bottom to top: integration of the T6P trace (2,661) using SIR of m/z 421.0, integration of the T6P trace (2,550) using MRM of m/z 78.6, 96.3, 138.7, 241.0 and 421.0, integration for each fragment ion m/z 78.6 (801), m/z 96.3 (868), m/z 138.7 (76), m/z 241.0 (404) and m/z 421.0 (392).



Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Analysis of *A. thaliana* plantlets following treatment. a–c, Phenotype analysis. a, Fresh weight of plantlets versus concentration of signalling precursors of T6P (1–4) and G6P precursors (14–17) in medium after three days (72 h) of uptake. Data are shown as mean \pm s.e.m.; $n = 3$. Each T6P precursor is shown (top) together with its G6P analogue (bottom). Visual appearance of a typical plantlet was analysed for a given concentration of precursors at the point of harvest. **b,** Phenotype of plants at the end of light treatments. Plants were allowed to take up compounds for 72 h and were then treated the next day with light treatments. Light treatments: GL, growth light irradiance $250 \mu\text{mol m}^{-2} \text{s}^{-1}$. UV 8 W and UV 100 W were growth light irradiance supplemented with UV light (365 nm). Daylight, part sun/part cloud; irradiance between $250 \mu\text{mol m}^{-2} \text{s}^{-1}$ under cloud and $1,440 \mu\text{mol m}^{-2} \text{s}^{-1}$ under full sun. Compounds were fed to the plants to a final concentration of 1 mM. Phenotype of plants fed with compound 3 at a reduced final concentration of 0.1 mM are shown in the right-hand panel. Scale, diameter of the plastic tube mouth = 10 mm. **c,** Typical *A. thaliana* phenotypes in the starch experiment. Plants were treated with a final medium concentration of 0.1 mM compound or water for 72 h and then exposed to 8-h, 8 W UV treatment. The plants were allowed to recover for another 24 h and were harvested at the end of the day and the starch content was measured. No significant phenotypic

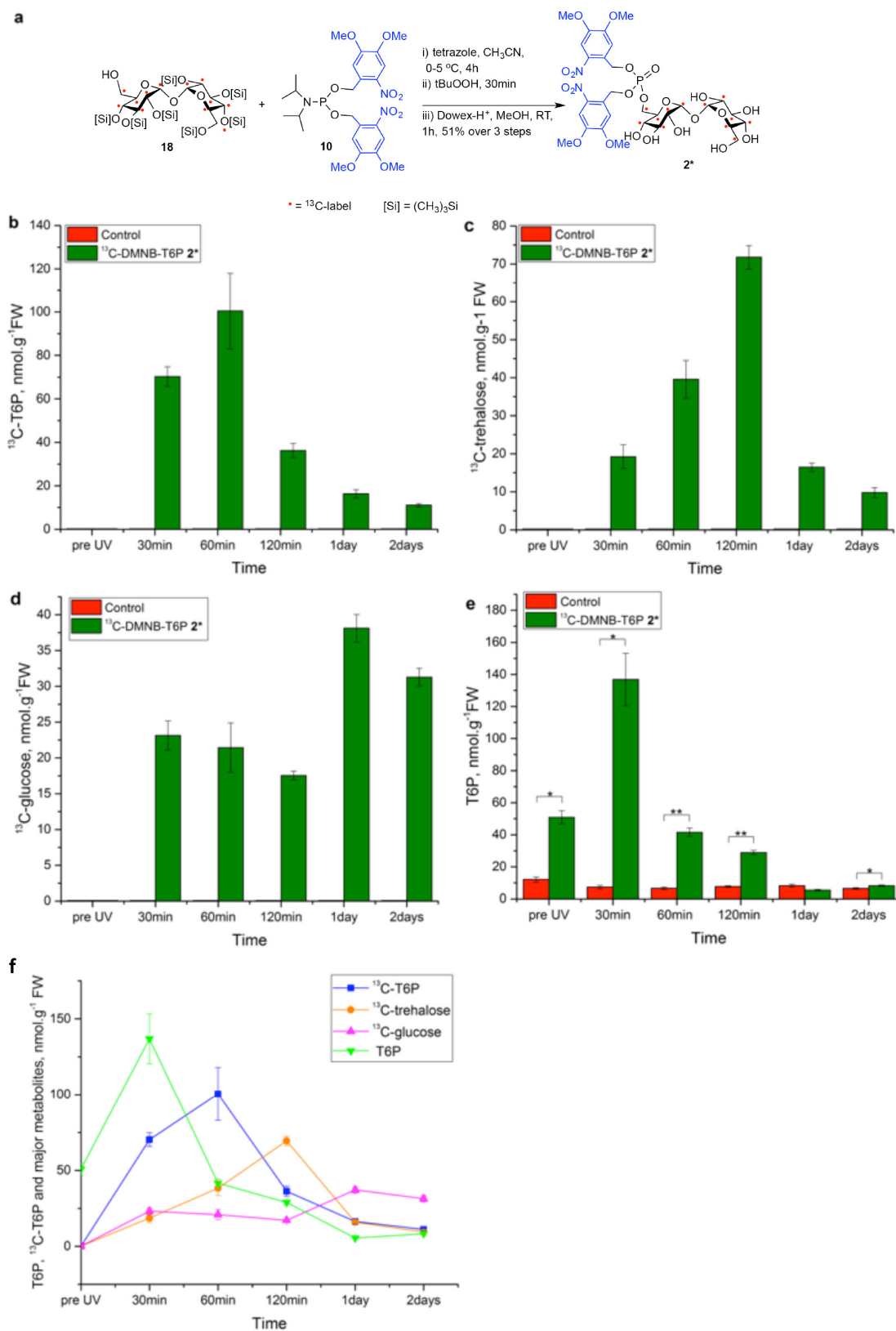
differences were observed between treatments. Water (left), 0.1 mM oNPE-T6P (compound 3) (middle), oNPE-G6P(1-OMe) (compound 16) (right). Scale, tube diameter = 10 mm. **d–g,** Biosynthetic effects of increasing T6P *in planta*. **d,** Starch level at the end of the day in UV-treated (8 W, $23 \mu\text{mol m}^{-2} \text{s}^{-1}$, 8 h) plants fed with compound 3 + UV is significantly higher than plants treated with water + UV ($n = 9$, data are shown as mean \pm s.e.m.). Samples for starch were taken 1 day after UV treatment. **e,** ADP-glucose pyrophosphorylase (AGPase) activity is increased in UV + compound 3 plants compared to compound 3 only, UV only, water only, UV + water-treated plants and plants treated with compound 16 ($n = 3$, data are shown as mean \pm s.e.m.). **f,** Starch synthesis rate in UV-treated ($20 \mu\text{mol m}^{-2} \text{s}^{-1}$, 8 h) plants treated with compound 3 (data are shown as mean \pm s.e.m., $n = 3$). **g,** Starch level at the beginning (data are shown as mean \pm s.e.m.; $n = 3$) and at the end (data are shown as mean \pm s.e.m.; $n = 4$) of the day in UV ($20 \mu\text{mol m}^{-2} \text{s}^{-1}$, 8 h) + water-treated (solid circles) and UV + compound 3-treated (empty circles) plants. *A. thaliana* used in **e–g** were grown at a light regime of 12 h day–12 h night, at $250 \mu\text{mol m}^{-2} \text{s}^{-1}$, and 23°C day/ 18°C night temperatures, treated with compounds 18 d after sowing, and exposed to UV light 72 h after addition of the compound. Asterisks in **d–f** denote significance according to ANOVA ($P = 0.002$). Asterisk in **g** denotes significance by one-way ANOVA (LSD 5% = 11.19).



Extended Data Figure 6 | Quantification of *in planta* metabolites.

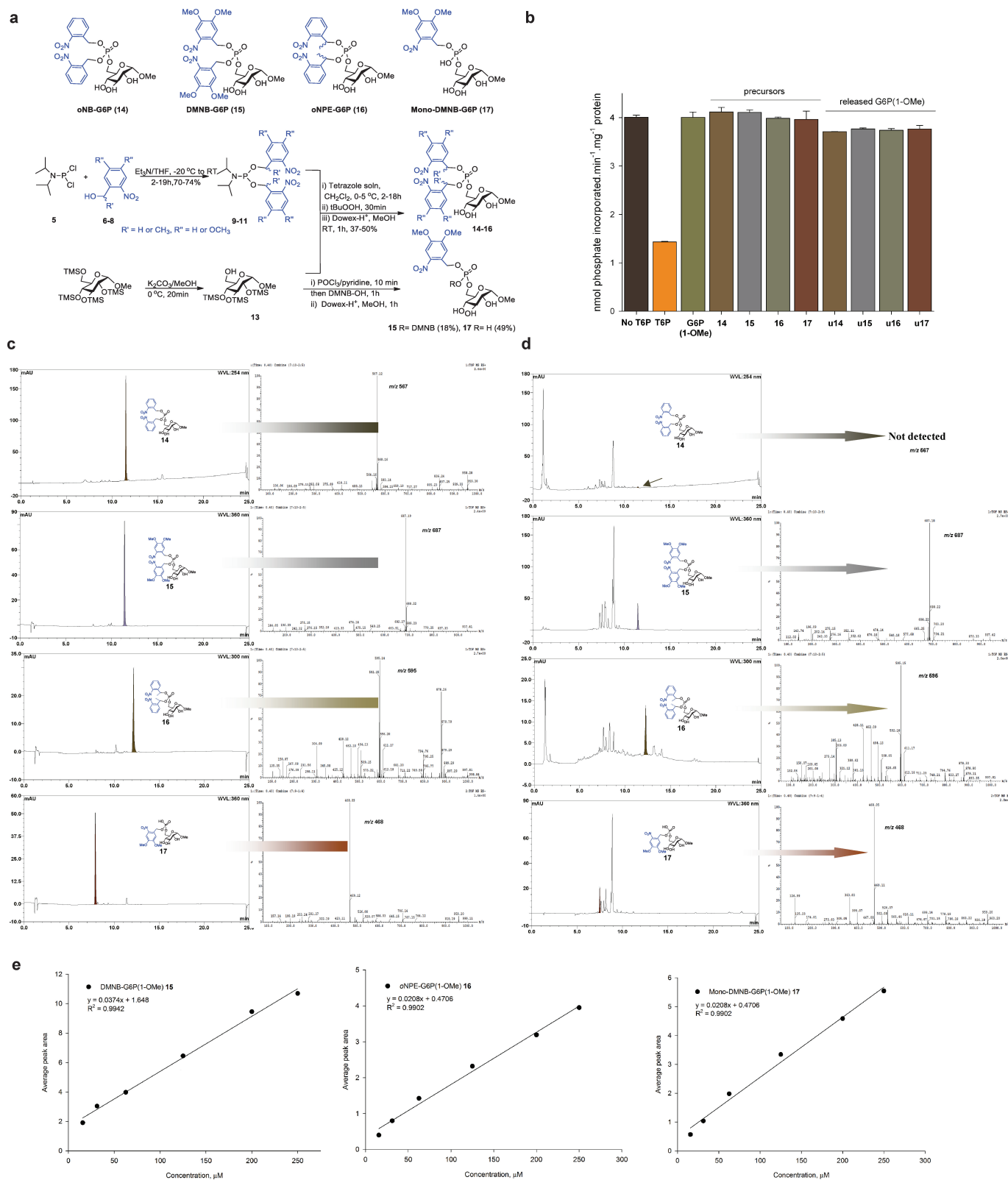
a, LC-MS chromatograms of trehalose, sucrose, glucose and fructose separation using a HILIC column, for details see Supplementary Information. **b**, Liquid chromatograms and peak areas of different concentrations of trehalose (100, 50, 25, 10 or 5 μM) and glucose (500, 250, 100, 50, 25, 10 or 5 μM). **c**, Calibration curves of the trehalose peak area against the concentrations (in μM) and glucose peak area against the concentrations (in μM). **d**, **e**, Same as Fig. 4, 7-day-old *A. thaliana* seedlings grown in liquid culture were treated with 1 mM of

compounds **3** or **2**, control seedlings were treated with water. Seedlings were left under growth lights to allow for uptake of the signalling precursors for 24 h, before exposure to $23 \mu\text{mol m}^{-2} \text{s}^{-1}$ UV for 2 h. Measurements were taken 1 day after uptake (pre-UV), 30, 60 and 120 min after initiation of UV treatment, and 1 and 2 d after initiation of UV treatment. See Fig. 4 for T6P content, trehalose content, sucrose content, glucose content; here fructose content (**d**) and Fresh weight biomass (**e**) are shown. * $P < 0.05$; ** $P < 0.01$ (Student's *t*-test). Data are shown as mean \pm s.e.m. ($n = 3$).

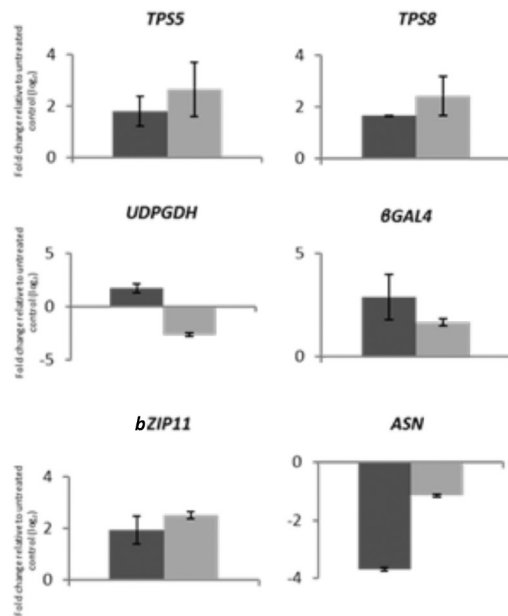
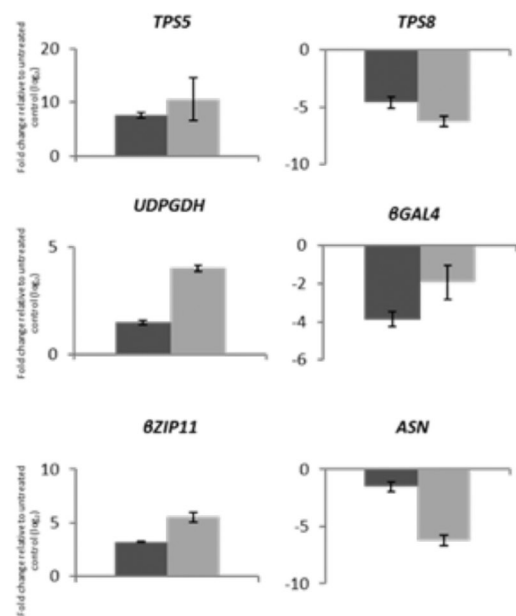
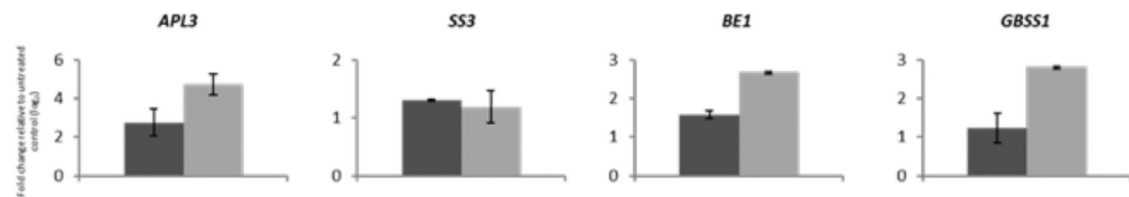
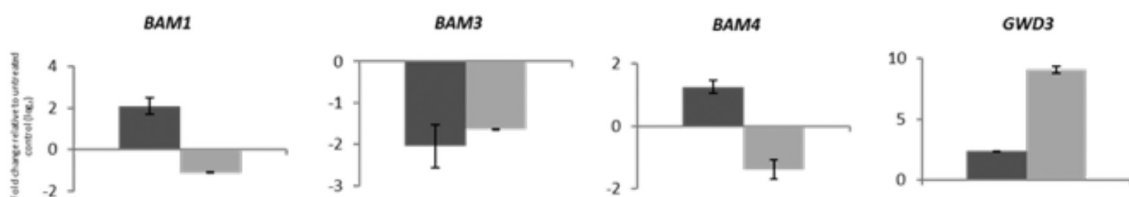


Extended Data Figure 7 | Dynamics of ¹³C-T6P, ¹³C-trehalose, ¹³C-glucose and T6P in *A. thaliana* treated with ¹³C-labelled precursor 2*. 1 mM of DMNB-¹³C-T6P (2*) was added to the growth medium of 7-day-old *A. thaliana* seedlings. The plants were left under growth light to allow uptake for 24 h and the uncaging was performed under 23 μmol m⁻² s⁻¹ UV for 2 h. Samples were harvested for analysis at different time points: pre-UV, 30, 60 and 120 min (after onset of

UV irradiation) and 1 and 2 days after UV irradiation. **a**, Synthesis of universally ¹³C-labelled 2* in essentially the same manner as for 2. **b**, Amount of ¹³C-T6P released over time in planta. **c**, Amount of ¹³C-trehalose accumulated. **d**, Amount of ¹³C-glucose accumulated. **e**, Amount of endogenous T6P. **f**, Overview of ¹³C tracking of T6P and metabolites. Data are shown as mean ± s.e.m.; *n* = 3; **P* < 0.05; ***P* < 0.01 (Student's *t*-test).



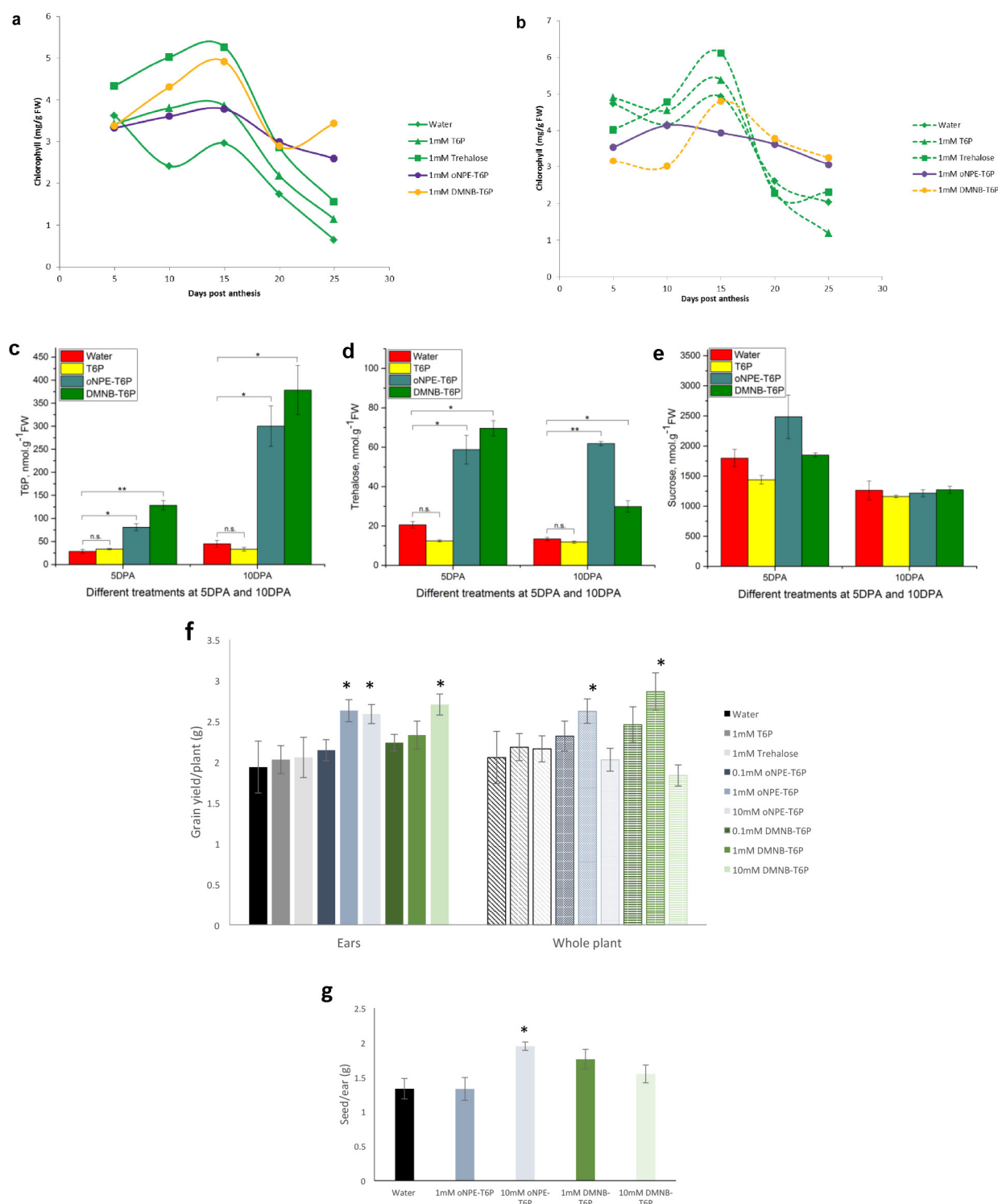
[M + Na]⁺ or [M – H][–], of pure G6P(1-Ome) precursors 14–17. For HPLC conditions see the Supplementary Information section 1. **d**, HPLC and mass spectrometry data, [M + Na]⁺ or [M – H][–], of uptaken G6P(1-Ome) precursors 14–17 *in planta*. **e**, Calibration curves for compounds 15, 16 and 17, respectively ($n = 2$ in all cases).

a 60 minutes UV treatment**b** One day post-UV treatment**c** Starch synthesis**d** Starch degradation

■ 1mM oNPE-T6P ■ 1mM DMNB-T6P $\bar{x} \pm \text{SEM}$

Extended Data Figure 9 | Transcript abundance of genes involved in starch synthesis and SnRK1 marker genes in response to caged T6P precursor application of 7-day-old *A. thaliana* seedlings in liquid culture. Seedlings were treated with a final concentration of 1 mM of compounds 2 or 3, uptake was allowed for 1 d under the growth lights, before treatment with $23 \mu\text{mol m}^{-2} \text{s}^{-1}$ UV light for 2 h to facilitate uncaging. **a, b**, Transcript fold change after 60 min of UV treatment (**a**) and 1 d after UV treatment (**b**) of SnRK1 marker genes. Marker genes normally downregulated by SnRK1: *TPS5*, *UDPGDH* (At3g29360) and *bZIP11*, and marker genes normally upregulated by SnRK1: *TPS8*, *BGAL*

and *ASN1*. **c**, Transcript fold change of starch synthesis genes after 60 min of UV treatment. Genes involved in starch synthesis: *APL3*, *SS3*, *BE1* and *GBSS1*. **d**, Transcript fold change after 60 min of UV treatment for starch degradation genes. Genes involved in starch degradation: *BAM1*, *BAM3*, *BAM4* and *GWD3*. Changes in transcripts for enzymes of degradation were more equivocal with *GWD3* increasing and *BAM* genes showing small changes or decreasing (*BAM3*). All data were normalized to a ubiquitin control. Data are shown as mean \pm s.e.m. of three independent samples.



Extended Data Figure 10 | Additional effects in wheat. **a**, Chlorophyll content of leaves after anthesis of ear treatments. **b**, Chlorophyll content of leaves after anthesis of whole-plant treatments. **c–e**, T6P release and metabolism in wheat. Developing wheat grain were treated with T6P, or compounds 2 or 3 (all 1 mM) 5 or 10 days after anthesis (DPA) and harvested 1 day later. **c**, Amount of T6P in wheat grains ($n=3$; data are shown as mean \pm s.e.m.). **d**, Trehalose ($n=3$; data are shown as mean \pm s.e.m.). **e**, Sucrose ($n=3$; data are shown as mean \pm s.e.m.).

* $P < 0.05$; ** $P < 0.01$ (Student's t -test). **f**, Dose response grain yield per plant to T6P precursors (0.1, 1 or 10 mM oNPE-T6P (3) or DMNB-T6P (2) and water, T6P and trehalose controls) sprayed to ears (5 ml) or to the whole plant (50 ml) at 5, 10, 15 and 20 days after anthesis. **g**, Grain yield per ear in response to a single time point spray (5 ml to the ear at 10 days after anthesis). * $P < 0.05$ (**f**, **g**) compared to water control (Student's t -test). Data are shown as mean \pm s.e.m. ($n=6$).

Integrin–YAP/TAZ–JNK cascade mediates atheroprotective effect of unidirectional shear flow

Li Wang^{1,2}, Jiang-Yun Luo^{1,2}, Bochuan Li^{3,4}, Xiao Yu Tian^{1,2}, Li-Jing Chen⁵, Yuhong Huang^{1,2}, Jian Liu¹, Dan Deng¹, Chi Wai Lau¹, Song Wan⁶, Ding Ai^{3,4}, King-Lun Kingston Mak², Ka Kui Tong⁷, Kin Ming Kwan⁷, Nanping Wang⁸, Jeng-Jiann Chiu⁵, Yi Zhu^{3,4} & Yu Huang^{1,2}

The *Yorkie* homologues YAP (Yes-associated protein) and TAZ (transcriptional coactivator with PDZ-binding motif, also known as WWTR1), effectors of the Hippo pathway, have been identified as mediators for mechanical stimuli¹. However, the role of YAP/TAZ in haemodynamics-induced mechanotransduction and pathogenesis of atherosclerosis remains unclear. Here we show that endothelial YAP/TAZ activity is regulated by different patterns of blood flow, and YAP/TAZ inhibition suppresses inflammation and retards atherosclerosis. Atheroprone-disturbed flow increases whereas atheroprotective unidirectional shear stress inhibits YAP/TAZ activity. Unidirectional shear stress activates integrin and promotes integrin–G α_{13} interaction, leading to RhoA inhibition and YAP phosphorylation and suppression. YAP/TAZ inhibition suppresses JNK signalling and downregulates pro-inflammatory genes expression, thereby reducing monocyte attachment and infiltration. *In vivo* endothelial-specific YAP overexpression exacerbates, while CRISPR/Cas9-mediated *Yap* knockdown in endothelium retards, plaque formation in *ApoE*^{−/−} mice. We also show several existing anti-atherosclerotic agents such as statins inhibit YAP/TAZ transactivation. On the other hand, simvastatin fails to suppress constitutively active YAP/TAZ-induced pro-inflammatory gene expression in endothelial cells, indicating that YAP/TAZ inhibition could contribute to the anti-inflammatory effect of simvastatin. Furthermore, activation of integrin by oral administration of MnCl₂ reduces plaque formation. Taken together, our results indicate that integrin–G α_{13} –RhoA–YAP pathway holds promise as a novel drug target against atherosclerosis.

Endothelial cells (ECs) are constantly exposed to mechanical forces generated by blood flow. Different shear forces induce distinct cellular responses. Disturbed flow is associated with vascular inflammation and focal distribution of atherosclerotic lesions, while steady unidirectional shear stress (USS) is anti-inflammatory and atheroprotective².

The Hippo pathway, a newly identified kinase cascade, is involved in organ size control and tumour suppression. Activation of this pathway leads to inhibition of downstream effectors YAP/TAZ by promoting their phosphorylation and cytoplasmic retention³. YAP/TAZ were reported as sensors for mechanical stimuli including matrix stiffness, stretch and cell density¹. However, the role of YAP/TAZ in haemodynamics-mediated signal transduction and atherosclerosis is still unclear.

Indirect evidence implies possible involvement of YAP/TAZ in atherogenesis. The well-characterized YAP/TAZ target genes (*CTGF* and *CYR61*) are highly expressed in human atherosclerotic lesions⁴. Lysophosphatidic acid, a major atherogenic factor, is the potent activator of YAP and TAZ⁵. Statins, the widely used anti-atherosclerotic drugs, were identified as the strongest YAP inhibitors

among 640 clinically used drugs⁶. However, direct evidence for YAP/TAZ activation in atherogenesis is still lacking.

We first found that mouse ECs express a higher level of YAP than other cells in aorta, indicating a possible role of YAP in maintaining

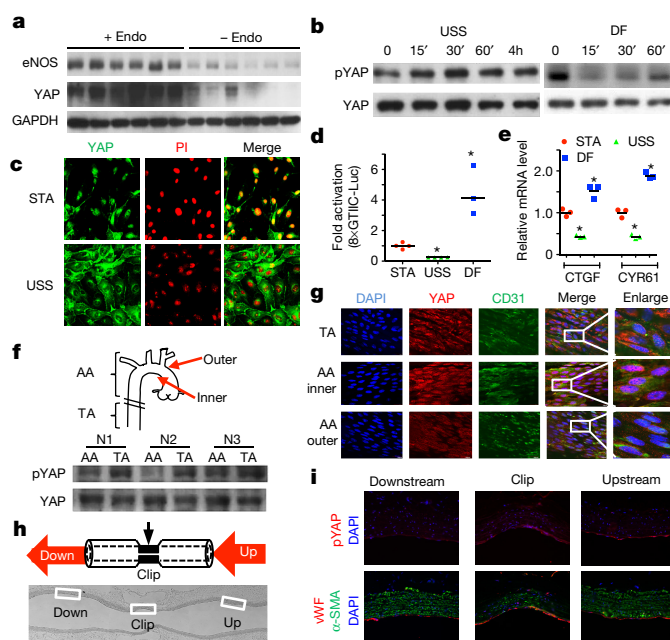


Figure 1 | Haemodynamics regulates YAP phosphorylation, subcellular localization, downstream gene expression and reporter gene activity in ECs. **a**, Immunoblotting showing YAP expression is higher in mouse aorta with endothelium (+ Endo) than that without endothelium (− Endo). **b**, USS promotes, while disturbed flow inhibits, YAP phosphorylation. **c**, USS promotes YAP nuclear exportation in HUVECs. YAP was visualized by immunostaining (green); nuclei were counterstained with propidium iodide (PI, red). **d**, **e**, USS suppresses while disturbed flow increases **(d)** 8×GTIC-luc reporter gene activity and **(e)** expression of YAP/TAZ target genes *CTGF* and *CYR61* ($n = 3$; compared with static (STA), $*P < 0.05$ by two-tailed unpaired *t*-test). **f**, Immunoblotting showing YAP phosphorylation level is higher in thoracic aorta (TA, straight) compared with aortic arch (AA, curved) from C57BL/6J mice. **g**, *En face* staining of YAP in mouse aorta showing increased YAP nuclear localization in inner curvature of the aortic arch compared with outer curvature and thoracic aorta ($n_{TA} = 6$, $n_{AA, inner} = 3$, $n_{AA, outer} = 3$). **h**, **i**, Immunostaining **(i)** of pYAP in rat abdominal aorta with surgical stenosis **(h)**, showing increased pYAP in the clipped region and decreased pYAP in the downstream region ($n = 3$). Representative images of three separate experiments are shown.

¹Institute of Vascular Medicine, Shenzhen Research Institute and Li Ka Shing Institute of Health Sciences, Chinese University of Hong Kong, Hong Kong, China. ²School of Biomedical Sciences, Chinese University of Hong Kong, Hong Kong, China. ³Collaborative Innovation Center of Tianjin for Medical Epigenetics, Tianjin Medical University, Tianjin 300070, China. ⁴Department of Physiology and Pathophysiology, Tianjin Medical University, Tianjin 300070, China. ⁵Institute of Cellular and System Medicine, National Health Research Institutes, Miaoli 35053, Taiwan. ⁶Department of Surgery, Chinese University of Hong Kong, Hong Kong, China. ⁷School of Life Sciences, Chinese University of Hong Kong, Hong Kong, China. ⁸Advanced Institute for Medical Sciences, Dalian Medical University, Dalian 116044, China.

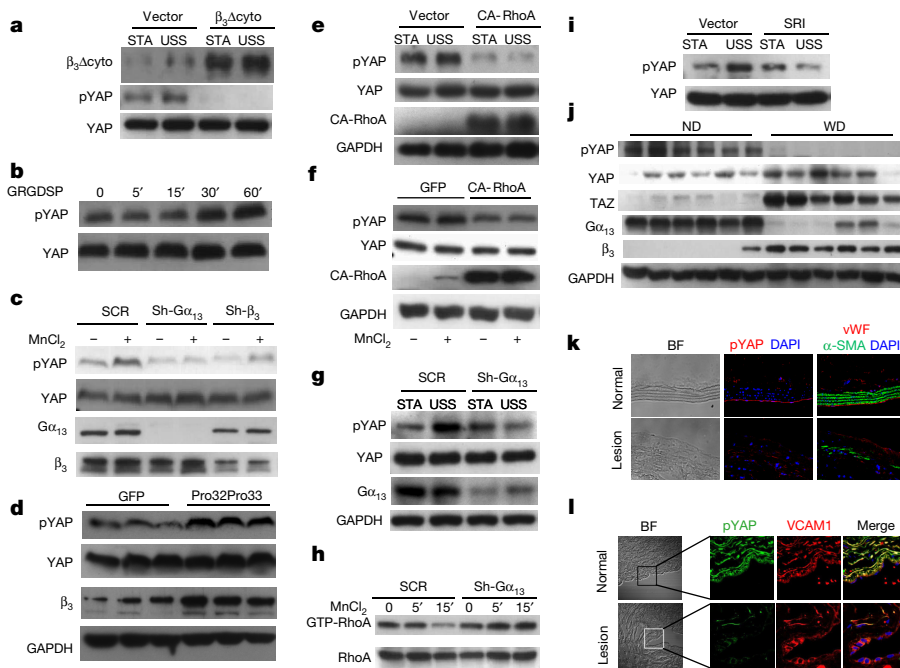


Figure 2 | Integrin inhibits YAP/TAZ activity through $G\alpha_{13}$ -mediated RhoA inhibition.

a, Cytoplasmic-domain-deleted integrin ($\beta_3\Delta\text{cyto}$) reverses USS-induced YAP phosphorylation in HUVECs. **b**, RGD-containing peptide GRGDSP ($100\mu\text{g ml}^{-1}$) induces YAP phosphorylation. **c**, Knockdown of $G\alpha_{13}$ or integrin β_3 attenuates MnCl_2 -induced (0.5 mM for 5 min) YAP phosphorylation. **d**, Integrin β_3 Pro32Pro33 mutation induces YAP phosphorylation. **e**, **f**, CA-RhoA suppresses YAP phosphorylation induced by (e) USS or (f) 0.5 mM MnCl_2 . **g**, **h**, $G\alpha_{13}$ knockdown reverses (g) USS-induced YAP phosphorylation and (h) MnCl_2 -induced GTP-RhoA suppression. **i**, $G\alpha_{13}$ inhibiting peptide SRI overexpression reverses USS-induced YAP phosphorylation. **j**, The expression of integrin β_3 , $G\alpha_{13}$, pYAP, YAP and TAZ in atherosclerotic aorta of $\text{ApoE}^{-/-}$ mice on Western diet (WD) or normal diet (ND) for 3 months. **k**, **l**, pYAP level in atherosclerotic lesions of (k) $\text{ApoE}^{-/-}$ mice ($n=5$) and (l) human ($n=5$). pYAP (red) (k) or green (l), vWF (red), VCAM1 (red) and $\alpha\text{-SMA}$ (green) were visualized by immunostaining; nuclei were counterstained with DAPI (blue). The representative images of at least three separate experiments are shown.

endothelium homeostasis (Fig. 1a). To investigate the impact of haemodynamics on YAP activity, YAP phosphorylation (Ser127, pYAP) in human umbilical vein ECs (HUVECs) subjected to USS (12 dyn cm^{-2}) or disturbed flow ($0.5 \pm 6\text{ dyn cm}^{-2}$, 1 Hz) was measured. Interestingly, USS inhibited, while disturbed flow activated, YAP activity. pYAP increased in HUVECs and human aortic ECs exposed to USS (Fig. 1b and Extended Data Fig. 1a). Accordingly, increased YAP/TAZ cytoplasmic retention was observed in HUVECs subjected to USS (Fig. 1c and Extended Data Fig. 1b, c). Congruently, USS suppressed transactivation activity of YAP/TAZ, indicated by reduced YAP/TAZ responsive luciferase ($8\times\text{GTIIC-luc}$) reporter gene activity and down-regulated expression of target genes (Fig. 1d, e). By contrast, disturbed flow reduced pYAP (Fig. 1b, and Extended Data Fig. 1d), enhanced YAP/TAZ reporter gene activity (Fig. 1d) and increased YAP/TAZ target gene expression (Fig. 1e and Extended Data Fig. 1e–g). To investigate the effect of haemodynamics on YAP activity *in vivo*, we determined YAP phosphorylation and nuclear localization in segments of mouse aorta and showed that pYAP level in aortic arch, an area exposed to disturbed flow, was lower than in thoracic aorta, an area exposed to USS (Fig. 1f). Consistently, in outer curvature of aortic arch and thoracic aorta, where blood flow is unidirectional, YAP was predominantly localized in the cytoplasm, while in the inner curvature of aortic arch, where blood flow is disturbed, YAP was mainly localized in the nuclei (Fig. 1g and Extended Data Fig. 1h). Rat abdominal aorta cross-clamping is a model used to generate different flow patterns *in vivo*⁷ (Fig. 1h). The constricted region, where unidirectional flow is accelerated, exhibited highest pYAP levels. Modest pYAP levels were detected in the upstream region where blood flow is unidirectional, while low pYAP was observed in the downstream region where blood flow is disturbed (Fig. 1i).

Integrin β_3 is a direct sensor for shear forces. The putative integrin agonists RGD-containing peptide (GRGDSP) or MnCl_2 can mimic the effect of USS⁸. To determine whether USS induces YAP phosphorylation through activating integrin, we examined USS-induced YAP phosphorylation in HUVECs transfected with loss-of-function mutation of integrin (with cytoplasmic domain deletion ($\beta_3\Delta\text{cyto}$))⁹. We found $\beta_3\Delta\text{cyto}$ overexpression abolished USS-induced YAP phosphorylation (Fig. 2a). Furthermore, treatment with GRGDSP or MnCl_2 increased pYAP in HUVECs (Fig. 2b and Extended Data Fig. 2a). In addition, GRGDSP suppressed YAP/TAZ target gene expression (Extended Data Fig. 2d). Congruently, MnCl_2 induced YAP/TAZ nuclear exportation (Extended Data Fig. 2b) and reduced YAP/TAZ reporter gene activity (Extended

Data Fig. 2c), whereas integrin β_3 knockdown reversed MnCl_2 -induced YAP phosphorylation (Fig. 2c). This evidence suggests that integrin activation directly induces YAP phosphorylation.

A previous study has suggested that flow-derived pulling force induces integrin activation by maintaining its extended conformation (ligand binding conformation)¹⁰. To test whether integrin β_3 in the extended conformation promotes YAP/TAZ phosphorylation, a Leu33Pro point mutation of integrin β_3 (Pro32Pro33 integrin) was constructed to mimic integrin β_3 activation¹¹. Indeed, Pro32Pro33 overexpression in HUVECs induced YAP phosphorylation (Fig. 2d), downregulated YAP/TAZ target gene expression (Extended Data Fig. 2e) and suppressed YAP/TAZ reporter gene activity (Extended Data Fig. 2f), indicating integrin β_3 mediates USS-induced YAP inhibition.

RhoA is one of the most important upstream activators of YAP/TAZ³. Integrin engagement and USS suppress RhoA activity^{8,12}. Therefore, we hypothesized that RhoA mediates integrin-induced YAP/TAZ suppression. As expected, basal and USS- or MnCl_2 -induced YAP phosphorylation was reduced in HUVECs transfected with constitutively active RhoA (Q63L) (CA-RhoA) (Fig. 2e, f).

G-protein subunit $G\alpha_{13}$ mediates integrin-induced RhoA suppression^{13–15}. Therefore, we investigated the effect of $G\alpha_{13}$ knockdown in USS- or MnCl_2 -induced RhoA inhibition and YAP phosphorylation. Neither USS nor MnCl_2 induced YAP phosphorylation when $G\alpha_{13}$ was silenced (Fig. 2c, g). Consistently, $G\alpha_{13}$ knockdown reduced MnCl_2 -induced YAP nuclear exportation (Extended Data Fig. 2g) and RhoA inhibition (Fig. 2h). Similarly, $G\alpha_{13}$ knockdown mitigated GRGDSP-induced suppression of YAP/TAZ target gene expression (Extended Data Fig. 2h).

Physical interaction between integrin β_3 and $G\alpha_{13}$ induces RhoA inhibition^{13,14}. To understand whether integrin β_3 and $G\alpha_{13}$ interaction mediates YAP phosphorylation, two myristoylated cell-permeable short peptides, mSRI and mP6, which mimic the interaction domain of $G\alpha_{13}$ and integrin β_3 , respectively, were used to selectively block association between $G\alpha_{13}$ and integrin β_3 (refs 13, 15). Similar to the effect of $G\alpha_{13}$ or integrin β_3 knockdown, mSRI or mP6 pretreatment abolished MnCl_2 -induced suppression of YAP/TAZ reporter gene activity and YAP phosphorylation in HUVECs (Extended Data Fig. 2c, i). Likewise, overexpression of SRI ($G\alpha_{13}$ blocking peptide) in HUVECs abolished USS-induced YAP phosphorylation (Fig. 2i).

Since haemodynamics is closely associated with pathogenesis of atherosclerosis, we compared the expression of YAP, pYAP, TAZ, $G\alpha_{13}$

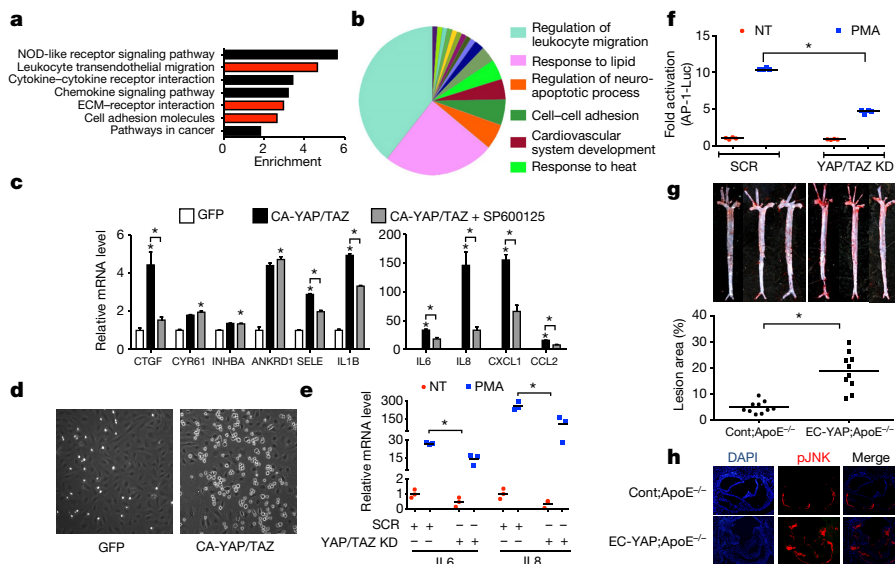


Figure 3 | YAP/TAZ activation induces adhesion molecule expression through increasing JNK activity. **a**, KEGG enrichment pathway analysis and **(b)** Gene Ontology (GO) enrichment analysis for mRNA profile in HUVECs transfected with CA-YAP/TAZ. **c**, JNK inhibitor SP600125 suppresses CA-YAP/TAZ-induced inflammatory gene expression in HUVECs ($n = 3$; $*P < 0.05$ by two-tailed unpaired t -test). **d**, CA-YAP/TAZ promotes monocyte attachment to HUVECs. **e**, **f**, YAP/TAZ knockdown reduces expression of **(e)** JNK target genes (*IL6* and *IL8*) and **(f)** AP-1 reporter gene activity induced by PMA ($n = 3$; $*P < 0.05$ by two-tailed unpaired t -test). **g**, **h**, EC-specific YAP overexpression promotes **(g)** atherosclerotic plaque formation, visualized by Oil Red O staining and **(h)** JNK activation, detected by immunostaining of pJNK (red) ($n = 5$, representative result is shown).

and integrin β_3 in aortas of *ApoE*^{-/-} mice with or without Western-diet-induced atherosclerosis. Our results showed downregulation of pYAP and α_{13} , and upregulation of TAZ, in aortas with atherosclerotic plaques (Fig. 2j). Consistent with a previous report¹⁶, we found integrin β_3 was highly expressed in mouse aortas with atherosclerosis, possibly because of compensatory response¹⁷. Immunofluorescence also showed that YAP phosphorylation was reduced in lesion area of *ApoE*^{-/-} mice and in human atherosclerotic aortas (Fig. 2k, l). Taken together, our results reveal that integrin activation promotes integrin- α_{13} association, which leads to RhoA suppression and subsequent YAP phosphorylation.

To explore the mechanism of YAP/TAZ activation in atherogenesis, we analysed messenger RNA (mRNA) profiles in HUVECs transfected with constitutively active YAP (S127A) and TAZ (S89A) (CA-YAP/TAZ). Four hundred and sixteen differentially expressed genes were identified by RNA sequencing (RNA-seq) ($P < 0.05$ and fold change cut-off > 1.5). DAVID KEGG enrichment analysis¹⁸ revealed six enriched pathways (Fig. 3a), including 'leukocyte transendothelial migration', 'ECM-receptor interaction' and 'cell adhesion molecules', etc. Gene Ontology enrichment for biological process analysed by GlueGo¹⁹ indicated YAP/TAZ is associated with regulation of leukocyte migration (Fig. 3b). Indeed, we observed more monocyte-endothelial adhesion associated with YAP/TAZ activation in HUVECs (Fig. 3d and Extended Data Fig. 3c). Moreover, several pro-inflammatory markers, such as *IL6*, *IL8* and *SELE*, were induced by YAP/TAZ activation (Fig. 3c and Extended Data Fig. 3a). Promoter reporter assay showed that CA-YAP/TAZ induced expression of adhesion molecules by enhancing their transcription (Extended Data Fig. 3b). However, deletion of the predicted TEAD binding sites, the known consensus DNA sequence for YAP-TEAD binding²⁰, in *CXCL1* and *SELE* promoters failed to reverse YAP/TAZ-induced reporter gene activity (data not shown), indicating other regulatory mechanisms might be involved. These results suggest that endothelial YAP/TAZ activation participates in the initiation of atherosclerosis by promoting monocyte adhesion.

JNK is critical in atherogenesis²¹. USS inhibits tumour-necrosis factor- α -induced JNK activation, while prolonged disturbed flow activates JNK^{22,23}. Our results showed that both USS and disturbed flow transiently increased phospho-JNK. However, in contrast to sustained JNK phosphorylation in HUVECs exposed to disturbed flow, prolonged USS suppressed JNK phosphorylation (Extended Data Fig. 3d, e). JNK effector activator protein (AP)-1 activity is reportedly increased by YAP/TAZ through JNK-YAP interaction²⁴⁻²⁶. We therefore hypothesized that YAP/TAZ promotes endothelial activation through enhancing JNK activity. Indeed, JNK inhibitor SP600125 suppressed YAP/TAZ-induced pro-inflammatory gene expression

(Fig. 3c). On the other hand, YAP/TAZ knockdown reduced basal and phorbol ester (PMA)-induced phospho-JNK, expression of JNK target genes *IL6* and *IL8* as well as AP-1 reporter gene activity²⁷ (Fig. 3e, f and Extended Data Fig. 3f). Dominant-negative YAP (YAP S94A) suppressed PMA-induced AP-1 reporter gene activity, whereas CA-YAP/TAZ enhanced AP-1 reporter gene activity (Extended Data Fig. 3g, h). To assess whether YAP activates JNK and accelerates atherosclerotic plaque formation *in vivo*, we generated EC-specific YAP overexpression mice on *ApoE*^{-/-} background (*Tie2*^{Cre/+}; *Yap*-*COE*^{tg/+}; *ApoE*^{-/-} (EC-Yap; *ApoE*^{-/-})) (Extended Data Fig. 4a, b). After 4 weeks of feeding on Western diet, EC-Yap; *ApoE*^{-/-} mice showed significantly increased plaque formation (Fig. 3g), accompanied by increased expression of p-JNK and macrophage marker Mac3 compared with control littermates (Cont; *ApoE*^{-/-}) (Fig. 3h and Extended Data Fig. 4c, d). Similar total cholesterol and triglyceride levels suggested that the atherogenic effect of endothelial YAP is unlikely to be related to lipid metabolism (Extended Data Fig. 4e, f).

To demonstrate that disturbed flow-associated atherosclerosis is mediated by endothelial YAP activation *in vivo*, *ApoE*^{-/-} mice received partial ligation surgery on the left carotid artery to develop disturbed flow-enhanced atherosclerosis. EC-specific *Yap* knockdown was achieved by using EC-enhanced AAV-mediated CRISPR/Cas9 (ref. 28) genome-editing system controlled by EC-specific *ICAM2* promoter. Immunohistochemistry and western blotting showed efficient *Yap* knockdown in ECs (Fig. 4a, b). Three weeks after surgery, severe plaques developed in control *ApoE*^{-/-} mice. However, mice with EC-specific *Yap* knockdown exhibited reduced plaque formation (Fig. 4c). Mice injected with adenovirus-mediated *Taz* short hairpin RNA (shRNA) also showed delayed atherogenesis (Extended Data Fig. 5a-c). Furthermore, oral administration of MnCl₂ reduced plaque formation in *ApoE*^{-/-} mice on Western diet for 12 weeks (Fig. 4d), without affecting lipid profile or superoxide dismutase activity (Extended Data Fig. 5h, i). Conversely, plaque formation increased in mice injected with AAV expressing CA-YAP/TAZ (Extended Data Fig. 5d-f). In summary, both gain- and loss-of-function experiments *in vivo* show the importance of YAP/TAZ activation in atherogenesis.

To examine whether existing anti-atherosclerotic drugs inhibit YAP/TAZ activity, we tested several agents (Extended Data Table 1). In addition to statins, which inhibit YAP/TAZ in tumour cells⁶, apelin, ApoA1 and niacin also suppressed YAP/TAZ activity (Fig. 4e). To understand whether YAP/TAZ suppression contributes to the anti-inflammatory effect of statins, we transfected HUVECs with CA-YAP/TAZ. Compared with HUVECs transfected with vector control, simvastatin failed to suppress expression of pro-inflammatory genes induced by CA-YAP/TAZ, suggesting YAP/TAZ inhibition might be

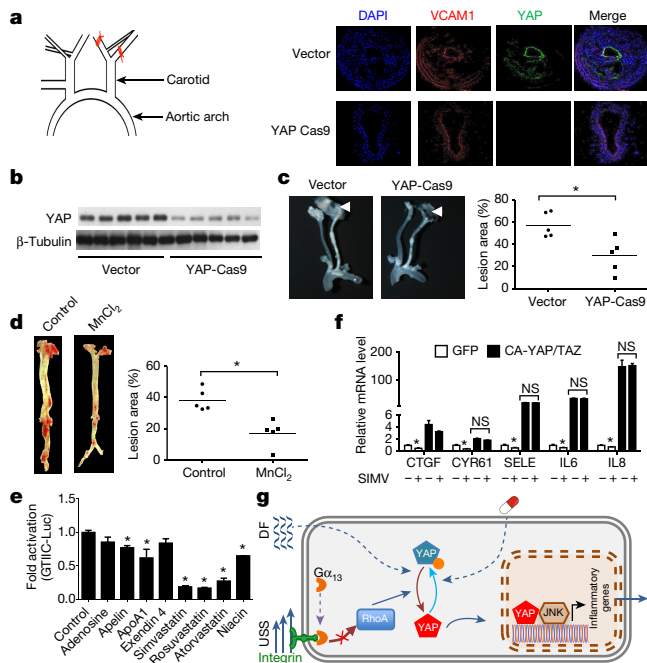


Figure 4 | Suppression of YAP/TAZ activity retards atherogenesis.

a, b, AAV-mediated CRISPR/Cas9 system specifically knocks down YAP level in endothelium of *ApoE*^{-/-} mice. Illustration (**a**, left) showing carotid partial ligation surgery in *ApoE*^{-/-} mice. YAP knockdown was confirmed by (**a**, right) immunostaining (YAP (green), Vcam1 (red), nuclei (blue)) ($n = 5$, representative result is shown) and (**b**) immunoblotting of YAP in aorta. **c**, EC-specific YAP knockdown reduces plaque formation in *ApoE*^{-/-} mice receiving carotid partial ligation (arrow) surgery. **d**, Oral administration of MnCl₂ decreases atherosclerotic plaque formation visualized by Oil Red O staining. **e**, YAP/TAZ reporter gene activity assay of anti-atherosclerotic agents showing statins produce the strongest inhibitory effect on YAP/TAZ activity ($n = 3$; * $P < 0.05$ by two-tailed unpaired *t*-test). **f**, Simvastatin suppresses expression of YAP/TAZ target genes while failing to reverse CA-YAP/TAZ-induced expression of pro-inflammatory genes ($n = 3$; * $P < 0.05$ by two-tailed unpaired *t*-test). NS, not significant. **g**, Illustration of the haemodynamics-regulated YAP/TAZ signalling in ECs.

involved in anti-inflammatory and anti-atherogenic effect of statins (Fig. 4f).

In summary, this study provides novel evidence showing that endothelial YAP/TAZ activation induced by atheroprone-disturbed flow promotes inflammation and atherogenesis by enhancing JNK activity, whereas the atheroprotective USS inhibits YAP/TAZ by modulating the integrin–Gα₁₃–RhoA pathway (Fig. 4g). Endothelial YAP/TAZ knockdown or MnCl₂ treatment delays atherogenesis, indicating YAP/TAZ could become a potential therapeutic target against atherosclerosis, as demonstrated by the YAP/TAZ-inhibitory effect of several anti-atherosclerotic drugs, especially statins.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 February; accepted 3 November 2016.

Published online 7 December; corrected online 21 December 2016

(see full-text HTML version for details).

- Halder, G., Dupont, S. & Piccolo, S. Transduction of mechanical and cytoskeletal cues by YAP and TAZ. *Nature Rev. Mol. Cell Biol.* **13**, 591–600 (2012).
- Cunningham, K. S. & Gotlieb, A. I. The role of shear stress in the pathogenesis of atherosclerosis. *Lab. Invest.* **85**, 9–23 (2005).
- Yu, F. X., Zhao, B. & Guan, K. L. Hippo pathway in organ size control, tissue homeostasis, and cancer. *Cell* **163**, 811–828 (2015).
- Chaquor, B. & Goppelt-Strube, M. Mechanical regulation of the Cyr61/CCN1 and CTGF/CCN2 proteins. *FEBS J.* **273**, 3639–3649 (2006).
- Yu, F. X. *et al.* Regulation of the Hippo–YAP pathway by G-protein-coupled receptor signaling. *Cell* **150**, 780–791 (2012).
- Sorrentino, G. *et al.* Metabolic control of YAP and TAZ by the mevalonate pathway. *Nature Cell Biol.* **16**, 357–366 (2014).

- Zhou, J. *et al.* Force-specific activation of Smad1/5 regulates vascular endothelial cell cycle progression in response to disturbed flow. *Proc. Natl Acad. Sci. USA* **109**, 7770–7775 (2012).
- Tzima, E., del Pozo, M. A., Shattil, S. J., Chien, S. & Schwartz, M. A. Activation of integrins in endothelial cells by fluid shear stress mediates Rho-dependent cytoskeletal alignment. *EMBO J.* **20**, 4639–4647 (2001).
- Ylänne, J. *et al.* Mutation of the cytoplasmic domain of the integrin β₃ subunit. Differential effects on cell spreading, recruitment to adhesion plaques, endocytosis, and phagocytosis. *J. Biol. Chem.* **270**, 9550–9557 (1995).
- Kim, C., Ye, F. & Ginsberg, M. H. Regulation of integrin activation. *Annu. Rev. Cell Dev. Biol.* **27**, 321–345 (2011).
- Vijayan, K. V. *et al.* Shear stress augments the enhanced adhesive phenotype of cells expressing the Pro33 isoform of integrin β₃. *FEBS Lett.* **540**, 41–46 (2003).
- Arthur, W. T., Petch, L. A. & Burridge, K. Integrin engagement suppresses RhoA activity via a c-Src-dependent mechanism. *Curr. Biol.* **10**, 719–722 (2000).
- Gong, H. *et al.* G protein subunit Gα₁₃ binds to integrin α₁₁β₃ and mediates integrin “outside-in” signaling. *Science* **327**, 340–343 (2010).
- Estevez, B., Shen, B. & Du, X. Targeting integrin and integrin signaling in treating thrombosis. *Arterioscler. Thromb. Vasc. Biol.* **35**, 24–29 (2015).
- Shen, B. *et al.* A directional switch of integrin signalling and a new anti-thrombotic strategy. *Nature* **503**, 131–135 (2013).
- Hoshiga, M., Alpers, C. E., Smith, L. L., Giachelli, C. M. & Schwartz, S. M. α_vβ₃ integrin expression in normal and atherosclerotic artery. *Circ. Res.* **77**, 1129–1135 (1995).
- Weng, S. *et al.* β₃ integrin deficiency promotes atherosclerosis and pulmonary inflammation in high-fat-fed, hyperlipidemic mice. *Proc. Natl Acad. Sci. USA* **100**, 6730–6735 (2003).
- Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).
- Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
- Zhao, B. *et al.* TEAD mediates YAP-dependent gene induction and growth control. *Genes Dev.* **22**, 1962–1971 (2008).
- Ricci, R. *et al.* Requirement of JNK2 for scavenger receptor A-mediated foam cell formation in atherogenesis. *Science* **306**, 1558–1561 (2004).
- Surapichat, J. *et al.* Fluid shear stress inhibits TNF-α activation of JNK but not ERK1/2 or p38 in human umbilical vein endothelial cells: inhibitory crosstalk among MAPK family members. *Proc. Natl Acad. Sci. USA* **98**, 6476–6481 (2001).
- Takabe, W. *et al.* Oscillatory shear stress induces mitochondrial superoxide production: implication of NADPH oxidase and c-Jun NH₂-terminal kinase signaling. *Antioxid. Redox Signal.* **15**, 1379–1388 (2011).
- Zanconato, F. *et al.* Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. *Nature Cell Biol.* **17**, 1218–1227 (2015).
- Liu, X. *et al.* Tead and AP-1 coordinate transcription and motility. *Cell Reports* **14**, 1169–1180 (2016).
- Ma, X. *et al.* Impaired Hippo signaling promotes Rho1-JNK-dependent growth. *Proc. Natl Acad. Sci. USA* **112**, 1065–1070 (2015).
- Hoffmann, E., Dittich-Breiholz, O., Holtmann, H. & Kracht, M. Multiple control of interleukin-8 gene expression. *J. Leukoc. Biol.* **72**, 847–855 (2002).
- Ran, F. A. *et al.* In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Chien for commenting on the manuscript, and C.-I. Lee and T.-E. Lin for conducting clip experiments. This study was supported by the Hong Kong Research Grants Council (CUHK2/CRF/12G), Natural Science Foundation of China (91339117, 81130002, 31430045), RGC (T12-402/13-N, C7055-14G, CUHK14105814), Croucher Foundation, CUHK Vice Chancellor's Discretionary Fund, Lui Che Woo Foundation, and the Ministry of Science and Technology, Taiwan (MOST104-2321-B-400-017, MOST104-2320-B-400-002-MY3).

Author Contributions L.W. designed the study, conducted most experiments, analysed the data and wrote the manuscript; J.Y.L. helped western blot, contributed ideas and prepared the manuscript. X.Y.T. helped revise the manuscript and provided suggestions for disturbed-flow-induced atherosclerosis; B.L. and D.A. generated EC-specific YAP transgenic mice and performed the *in vivo* study. Y.H.H. and D.D. performed plasmid construction. L.J.C., J.L. and C.W.L. performed immunohistochemistry and the carotid artery partial ligation surgery. J.J.C. and S.W. helped with atherosclerotic samples and contributed to data analysis. K.L.M., K.K.T. and K.M.K. helped in animal studies. J.J.C. and N.W. provided constructive suggestions in experimental design and helped revise the manuscript. Y.Z. and Y.H. are the leading principal investigators who directed the study and data analysis, and prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.H. (yu-huang@cuhk.edu.hk) or Y.Z. (zhuyi@tmu.edu.cn).

Reviewer Information Nature thanks P. F. Davies, G. Halder and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Antibodies. The antibodies used for western blotting included anti-YAP/TAZ (1:1,000; 8418; Cell Signaling Technology, USA), anti-YAP (1:1,000; Cell Signaling Technology, USA), anti-pYAP (1:1,000; Ser 127, 4911S; Cell Signaling Technology, USA), anti-TAZ (1:1,000; ab84927; Abcam, UK), anti-JNK (1:1,000; 9252h; Cell Signaling Technology, USA), anti-pJNK (1:1,000; 9255; Cell Signaling Technology, USA), anti-CTGF (1:1,000; ab6992; Abcam, UK), anti-G α_{13} (1:1,000; ab128900; Abcam, UK), anti-integrin β_3 (1:1,000; 4702; Cell Signaling Technology, USA), anti-RhoA (1:1,000; ab54835; Abcam, UK) and anti-eNOS (1:1,000; BD Biosciences, USA).

The antibodies used for immunostaining included anti-pYAP (1:100; Ser 127, 4911S; Cell Signaling Technology, USA), anti-YAP (1:100; Cell Signaling Technology, USA) and anti-pJNK (1:100; 9255; Cell Signaling Technology, USA).

Quantitative real-time PCR. RNA was extracted by using TRIzol Reagent (Thermo) according to the manufacturer's protocol. cDNA was synthesized using a High-Capacity cDNA Reverse Transcription Kit (Thermo). Quantitative PCR was performed using SYBR Select (Thermo) following the manufacturer's protocol. *GAPDH* was used as the internal control. Primers used for quantitative real-time PCR were included in Supplementary Table 1.

Western blotting. Cells or tissues were homogenized in cold RIPA lysis buffer supplemented with cOmplete Protease Inhibitors cocktail and phosphatase inhibitor (Roche). The protein concentration was determined using Bradford Assay (Bio-Rad). Ten micrograms of protein were resolved by SDS-polyacrylamide gel electrophoresis and transferred to the PVDF membrane (Bio-Rad). Target protein was detected using specific primary antibody. Bound antibodies were detected by horseradish-peroxidase-conjugated secondary antibody and visualized by enhanced chemiluminescence (Cell Signaling Technology). Experiments were repeated three times and the target protein level was quantified by ImageJ and normalized to internal control (or pYAP was normalized to total YAP) (Extended Data Figs 6 and 7). Original western blot scans are included in Supplementary Fig. 1.

Cell culture. HUVECs and human aortic ECs were purchased from Lonza (EGM, Clonetics, Lonza, Walkersville, Maryland, USA). Lonza guarantees that the cells express CD31/105, von Willebrand Factor VIII, and are positive for acetylated low-density lipoprotein uptake. We did not test for mycoplasma contamination during the experiments. HUVECs were maintained in EGM supplemented with EGS and FBS at 37 °C in an incubator with 95% humidified air and 5% CO₂ and passaged every 3 days. Cells within seven passages were used for the *in vitro* study.

GST-RBD pull-down for active RhoA detection. GST-RBD recombinant protein was purified from BL21 (DE3) *Escherichia coli* and affinity conjugated to glutathione sepharose beads (Pharmacia). For GST affinity pull-down, 10⁷ cells were lysed in 1 ml Weak Lysis Buffer (Beyotime) supplemented with protease inhibitors (Roche). Cell lysates were centrifuged at 15,000 g at 4 °C for 20 min to remove cell debris. Cell lysates were incubated in sepharose beads conjugated with 1 μ g GST-RBD and incubated at 4 °C for 2 h with constant agitation, and precipitated by centrifugation at 1,000 r.p.m. for 10 min. After three washes, beads were collected by centrifugation and boiled in 2 \times SDS loading buffer for 5 min. The active RhoA was determined by western blotting.

Experimental animals. Animals were supplied by the University Laboratory Animal Services Centre and their use approved by the Ethical Committee of Animal Research (CUHK). The animals used in the present study included Sprague-Dawley rats, apolipoprotein E deficient (*ApoE*^{-/-}) mice and EC-specific YAP overexpression transgenic mice. Male mice or rats were used in all *in vivo* studies. The animals were kept at a constant temperature (21 \pm 1 °C) under 12/12-h light/dark cycle and had free access to water and standard chow unless specified.

Construction of EC-specific YAP overexpression mice. CAG loxp-stop-loxp-Yap mice were generated in a C57BL/6 background in Model Animal Research Center (Nanjing, China). Yap-COE mice were crossed with *ApoE*^{-/-} mice and then Tie-2-Cre^{+/+} mice. The 6-week-old *ApoE*^{-/-};Yap-COE;Tie-2-Cre^{+/+} and *ApoE*^{-/-};Yap-COE;Tie-2-Cre^{+/+} mice were bred and housed in temperature-controlled cages under a 12/12-h light/dark cycle with free access to water in Tianjin Medical University Animal Center. Study protocols and the use of animals were approved by the Institutional Animal Care and Use Committee of Tianjin Medical University (Tianjin, China). The mice were fed a Western diet (Research Diets, D12109) containing 40 kcal% fat, 1.25% cholesterol and 0.5% cholic acid for 4 weeks before being killed. Aortas were isolated to assess lesion formation and distribution by Oil Red O staining. Aortic roots were stained for pJNK, α -SMA and macrophages.

En face staining. Mouse aortas were fixed with 4% paraformaldehyde for 15 min. After permeabilization/blocking in 0.05% Triton X-100 (in PBS) and 1% BSA and for 0.5 h at room temperature, aortas were incubated at 4 °C overnight in incubation buffer containing 1% BSA and the primary antibody including YAP1 (Abcam, ab52771), CD31 (Abcam, ab24590). After being washed in PBS three times, aortas

were incubated with Alexa-Fluor 488-, Alexa-Fluor 594-conjugated secondary antibodies (ZSGB-BIO, Beijing) for 1 h at room temperature. The fluorescent signal was detected by a Leica confocal laser scanning microscopy.

Disturbed flow *in vivo*. Stenosis of the abdominal aorta of rats was induced using a U-shaped titanium clip, as described^{29,30}. Briefly, after anaesthetization with isoflurane, the rat was laid supine and a lower midline abdomen incision was made; the part of the intestine was gently lifted out of the abdominal cavity and kept moist with saline throughout the surgical procedure. The aorta, left and right common iliac artery were exposed and the accompanying vein was carefully separated. The clip was held with a pair of forceps and placed around the isolated segment (1 cm from the arterial bifurcation) to partly constrict the abdominal aorta. The extent of clipping was controlled by placing a stopper of given size between the two arms of the forceps. Two weeks later, the rat was euthanized by intoxication with 100% carbon dioxide, and the aorta was perfusion-fixed with 4% (w/v) paraformaldehyde at 120 mm Hg. The fixed aorta was embedded in paraffin blocks for immunohistochemical staining.

Partial ligation of carotid artery was generated as described before³¹. Briefly, *ApoE*^{-/-} mice were anaesthetized by intraperitoneal injection of xylazine (10 mg/kg) and ketamine (80 mg/kg) mixture. A ventral midline incision (4–5 mm) was made in the neck. Left carotid artery was exposed by ventral midline incision (4–5 mm) in the neck. Left external carotid, internal carotid and occipital arteries were ligated, while the superior thyroid artery was left intact. Mice were monitored until recovery in a chamber on a heating pad after surgery and fed the Western diet immediately after surgery until killed.

Immunohistochemical staining. Immunohistochemical staining was performed on serial sections (5 μ m thick) of paraffin-embedded rat abdominal aortas and *ApoE*^{-/-} mouse aortas using pYAP (Cell Signaling), EC- and SMC-specific markers (that is, vWF and α -SMA, respectively) (Merck Millipore). Briefly, the sections were de-waxed in xylene, rehydrated in descending grades of alcohol and permeabilized by incubating for 10 min in sodium citrate for 10 min at 95 °C. Sections were cooled down to room temperature and blocked with blocking reagent (Merck Millipore) for 30 min. One section was incubated with antibody against pYAP (1:100) overnight at 4 °C, followed by Alexa-Fluor 594-conjugated goat-anti-rabbit IgG (1:1,000; Invitrogen) secondary antibody in blocking reagent for 1 h at room temperature. The secondary section was incubated with antibodies against vWF and α -SMA (1:100 each) overnight at 4 °C, followed by Alexa-Fluor 594-conjugated goat-anti-rabbit IgG and Alexa-Fluor 488-conjugated goat-anti-mouse IgG (1:1,000; Invitrogen) secondary antibodies in blocking reagent for 1 h at room temperature. Nuclei were co-stained by DAPI (Invitrogen) in PBS for 5 min. The sections were spin-dried and mounted with ProLong Gold (Invitrogen) on glass coverslips. Images were acquired and analysed using a Zeiss fluorescence microscope with Axiovision image analysis software.

Oral administration of MnCl₂ in *ApoE*^{-/-} mice. *ApoE*^{-/-} mice (male, 12 weeks old) were fed a Western diet, and MnCl₂ was administered through voluntary water consumption. Water consumption rate was predetermined by monitoring the volume of water remained. MnCl₂ was supplemented to drinking water to achieve 5 mg/kg body weight. Mice body weight and water consumption were adjusted weekly to adapt to the change of body weight and water consumption. After feeding on the Western diet for 3 months, the mice were killed and the atherosclerotic plaque formation was determined by Oil Red O staining.

Oil Red O staining for atherosclerotic plaques in mouse aorta. The *ApoE*^{-/-} mice were killed by CO₂ asphyxiation. Mouse aortas were dissected in cold PBS and cut open to expose the atherosclerotic plaques. After fixation in 4% formaldehyde for 16 h at 4 °C, the tissues were first rinsed in water for 10 min and then in 60% isopropanol. The aortas were stained with Oil Red O for 15 min with gentle shaking, and rinsed again in 60% isopropanol and then in water for three rinses. The samples were fixed on the cover slides with the endothelial surface facing upwards. The images were recorded using an HP Scanjet G4050. The plaque areas were determined using National Institutes of Health ImageJ software and calculated by expressing the plaque area relative to the total vascular area.

Human aortic specimens. The experiments were approved by the Hospital Human Subjects Review Committee (IRB approval number TSGHIRB 2-103-05-132) of Tri-Service General Hospital in Taipei and were conducted under the guidelines established by the Ethics Review Board of National Health Research Institutes, Taiwan. Written informed consent was obtained from all individuals. Human aortic tissue specimens were from patients with acute type-A aortic dissection. These samples were collected during emergency aortic surgery. The diseased segments of aorta (that is, dissecting aortic aneurysm) in these patients were all resected and replaced by an artificial inter-position graft. Specimens were fixed in paraformaldehyde, paraffin-embedded and cut into 5 μ m sections. YAP Ser127 phosphorylation was determined by immunofluorescence imaging.

RNA-sequencing. HUVECs were transfected with pWCXI-Flag-YAP-S127A (a gift from K. Guan, Addgene 33092) and 3 \times Flag pCMV5-TOPO TAZ (S89A)

(a gift from J. Wrana, Addgene 24815) or pEGFP-N1 by Neon transfection system (Invitrogen, USA)^{32,33}. Four hours after transfection, cells were harvested and RNA was extracted using RNeasy Mini Kit (Qiagen, Germany). The extracted RNA samples were sent to Beijing Genomics Institute (BGI) for RNA-sequencing analysis. $P < 0.05$ and fold change > 1.5 was used as a threshold for different regulated genes. DAVID tools were used for the pathways enrichment analysis and GlueGo was used for the Gene Ontology analysis.

Haemodynamics study *in vitro*. Ibidi flow system (IBIDI, Germany) was used to generate USS and disturbed flow (12 dyn cm^{-2} for USS and $0.5 \pm 6 \text{ dyn cm}^{-2}$, 1 Hz for disturbed flow). μ -slide I 0.4 Luers (IBIDI, LLC) was used for immunofluorescence studies. The slide was coated with $50 \mu\text{g/ml}$ fibronectin for 24 h. Seven thousand HUVECs were seeded onto the slide. After cells were adapted to medium containing 2% FBS (10% fatty acid free BSA for disturbed flow) for 6 h, the slides were mounted onto the Ibidi flow system. For immunostaining of USS-induced YAP/TAZ nuclear exportation, cells were subjected to USS for 6 h. For western blotting and reverse transcription real-time PCR analysis, the μ -slides were replaced with a custom-built flow chamber, which could accommodate more cells. Glass slides ($75 \text{ mm} \times 38 \text{ mm}$; Corning) were coated with fibronectin ($50 \mu\text{g/ml}$). HUVECs were seeded on slides and allowed to attach on the bottom for 16 h. For USS, the medium was replaced with EGM supplemented with 2% FBS for 6 h. For disturbed flow, cells were incubated in EGM supplemented with 10% fatty-acid-free BSA (Sigma). The slides were mounted onto the flow chamber and connected to the Ibidi flow system. The cells were then subjected to USS or disturbed flow. For USS-induced YAP phosphorylation, 15 min of shear force was applied unless otherwise noted. For USS-induced YAP translocation, 6 h of shear force was applied. For reverse transcription real-time PCR analysis, 4 h of shear stress was sufficient to inhibit the expression of YAP/TAZ target genes. For reporter gene assay, 48 h of shear forces were applied to HUVECs.

Plasmid construction. To construct the reporter plasmids for adhesion molecules, human genomic DNA was purified from HUVECs using a Universal Genomic DNA Extraction Kit Ver 3.0 (Takara, Japan). The promoters of *ICAM1*, *SELE*, *CCL2* and *CXCL1* were PCR amplified from human genomic DNA using the primers listed in Supplementary Table 1. A 2.1 kb fragment (-1784 to $+328$) from the *ICAM1* promoter, a 2.2 kb fragment (-1807 to $+475$) from the *SELE* promoter, a 4 kb fragment (-3992 to $+73$) from *CCL2* promoter and a 1.3 kb fragment (-1256 to $+84$) from *CXCL1* promoter were amplified. The PCR products were gel purified by gel extraction kit (Takara, Japan) and digested with restriction enzymes. The digested fragments were gel purified and ligated to pGL3 reporter plasmid digested by corresponding restriction enzymes. The ligation products were then heat inactivated at 65°C for 15 min and transformed into the DH5 α competent cells.

The Pro32Pro33 integrin was derived from pcDNA3.1-beta-3 (a gift from T. Springer, Addgene plasmid 27289) by point mutation³⁴.

Primers used for plasmids construction were included in Supplementary Table 1.

Adenovirus production. To generate the adenovirus shuttle vector pShuttle-U6, the U6 promoter and 1.9 kb stuffer sequence was excised from pLKO.1 (a gift from D. Root, Addgene plasmid 10878) with NotI/XhoI and ligated into pShuttle plasmid pre-digested with restriction enzymes accordingly. Short hairpin RNA targeting mouse *Taz* was generated using a protocol similar to pLKO.1 shRNA plasmids (Addgene) construction protocol. *Taz* shRNA sequence, TRCN0000095951, which was validated by Mission shRNA (Sigma Aldrich), was used to generate shuttle plasmids for *Taz* shRNA.

Recombinant adenovirus was generated using the AdEasy system³⁵. Briefly, pShuttle-U6 vector containing shRNA was digested with PmeI and co-transformed with adenoviral backbone plasmid pAdEasy-1 for homologous recombination in *E. coli* BJ5183 cells. Positive recombinants were linearized by PacI digestion and transfected into HEK-293A cells for virus packaging. The medium and cells were collected until the cytopathic effect was apparent. After three cycles of freeze and thaw to release the virus, the cell debris was removed by centrifugation at 3,000 r.p.m. for 15 min. The virus-containing supernatant was collected by PEG precipitation, followed by dialysis against saline with 100K MWCO dialysis tubing (Spectrum Labs).

Lentivirus production. Lentiviral shuttle plasmids for *YAP* (TRCN0000300325), *TAZ* (TRCN0000370007), $G\alpha_{13}$ (TRCN0000036885) and *ITGB3* (TRCN0000003236) shRNA were purchased from Sigma. Plasmid cocktail containing 1 μg of resultant shuttle plasmid, 750 ng of psPAX2 packaging plasmid and 250 ng of pMD2.G envelope plasmid were co-transfected to HEK-293FT cells. The medium was changed 15 h after transfection; 48 and 72 h after transfection, the medium containing the lentiviral particles was harvested then passed through 0.45 μm filters to remove cell debris. The virus was precipitated with PEG and suspended in PBS containing 4% sucrose. The lentiviral solutions were then aliquoted to vials and stored at -80°C .

Construction of AAV shuttling plasmid for CA-YAP/TAZ overexpression.

YAP1 S127A was amplified from pWCXIH-Flag-YAP-S127A (a gift from K. Guan, Addgene 33092) and ligated to pAAV-MCS (Stratagene) to generate the pAAV-YAP1 S127A shuttle plasmid. A similar strategy was used to generate the pAAV-TAZ S89A from 3 \times Flag pCMV5-TOPO TAZ (S89A) (a gift from J. Wrana, Addgene 24815).

Construction of endothelial specific AAV-mediated CRISPR/Cas9 shuttle plasmid. pX601-AAV-CMV: NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA (a gift from F. Zhang, Addgene plasmid 61591) was used to generate the EC-specific Cas9 for *Yap* *in vivo* genome editing²⁸. Three sgRNA sequences for *Yap* were predicted by CCTop (CRISPR/Cas9 target online predictor)³⁶. *ICAM2* endothelium-specific promoter from human was synthesized by GenScript and replaced the CMV promoter in pX601-AAV-CMV¹⁴.

Primers used for sgRNA were included in Supplementary Table 1.

Endothelial enhanced AAV packaging. The shuttle plasmids were co-transfected into HEK-293T with endothelial enhanced RGDLRVS-AAV-*cap* plasmid (provided by O. J. Müller, Universität Heidelberg, Germany) and pHelper plasmid (Stratagene)³⁷. After co-transfection for 72 h, the AAV viral particles were isolated according to the protocol reported in ref. 38. Briefly, the cells were harvested and re-suspended in $1\times$ restore buffer and the nuclei were extracted by homogenization. Viral particles were extracted by using nuclear lysis buffer. The viral particles were purified by PEG concentration, followed by dialysis against saline with 100K MWCO dialysis tubing (Spectrum Labs) to remove impurities, and concentrated. The viral titration was determined by qPCR and adjusted to 10^{10} plaque-forming units per ml in PBS containing 4% sucrose.

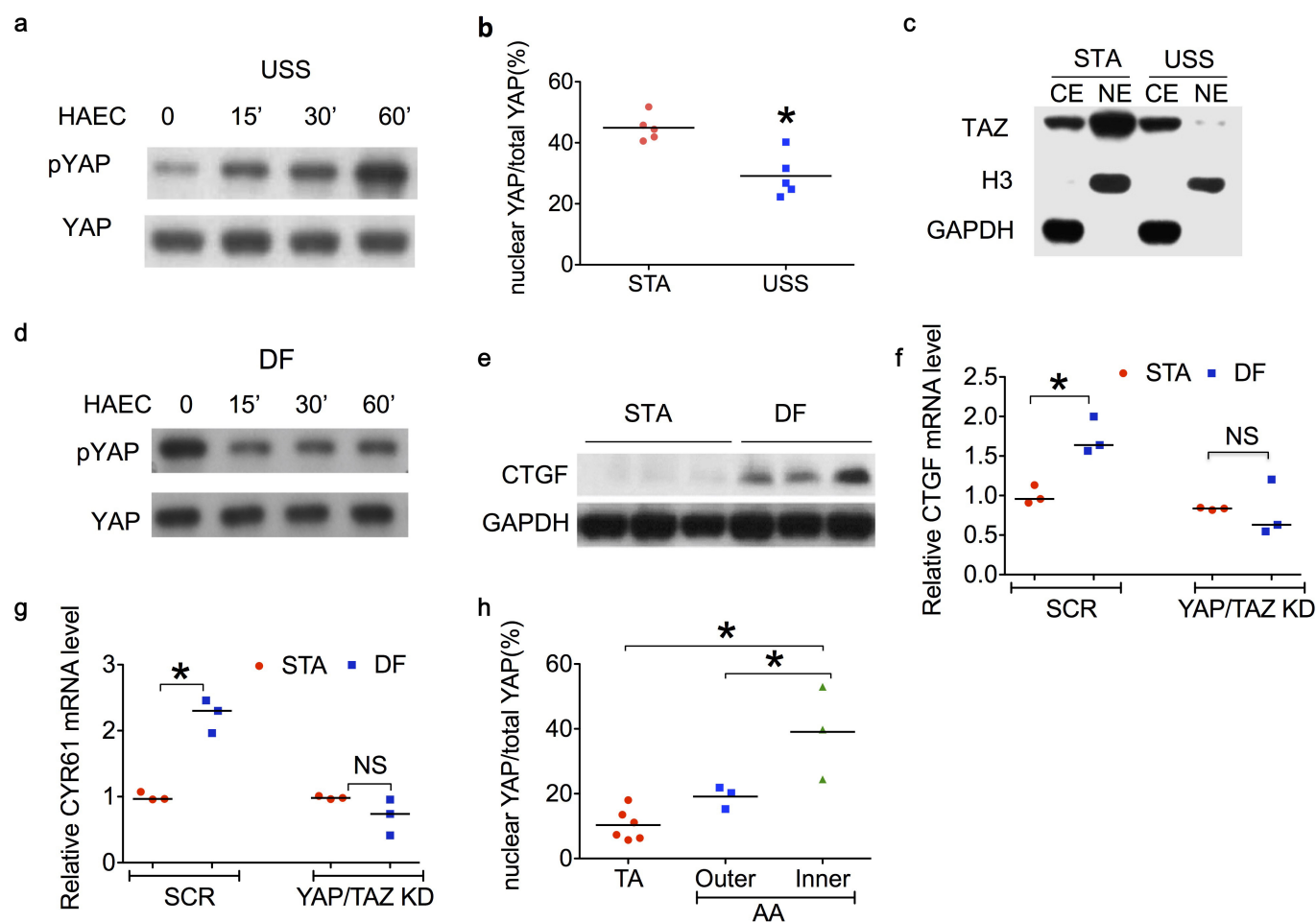
Virus administration. For adenovirus-mediated *Taz* shRNA, viruses (10^9 plaque-forming units) were administered to *ApoE*^{-/-} mice (male, 12 weeks old) that had been fed on Western diet (Research Diets) for 4 weeks, through tail vein injection. The mice were then fed on Western diet for 2 more months. The atherosclerotic plaque formation was visualized by Oil Red O staining.

For AAV-mediated CA-YAP/TAZ overexpression and YAP-Cas9, the viruses (10^9 plaque-forming units) were administered to *ApoE*^{-/-} mice (male, 12 weeks old) through tail vein injection before feeding on Western diet or receiving the carotid partial ligation surgery.

Statistical analysis. Statistics analyses were performed using GraphPad Prism 5.0. The sample sizes were not predetermined by statistical methods. The samples were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. At least three independent experiments were performed for all biochemical experiments and the representative images were shown. Results represent mean \pm s.e.m. Student's *t*-test (unpaired two-tailed) was used in the analysis. No samples, mice or data points were excluded from the reported analysis. Levels of probabilities less than 0.05 were regarded as significant.

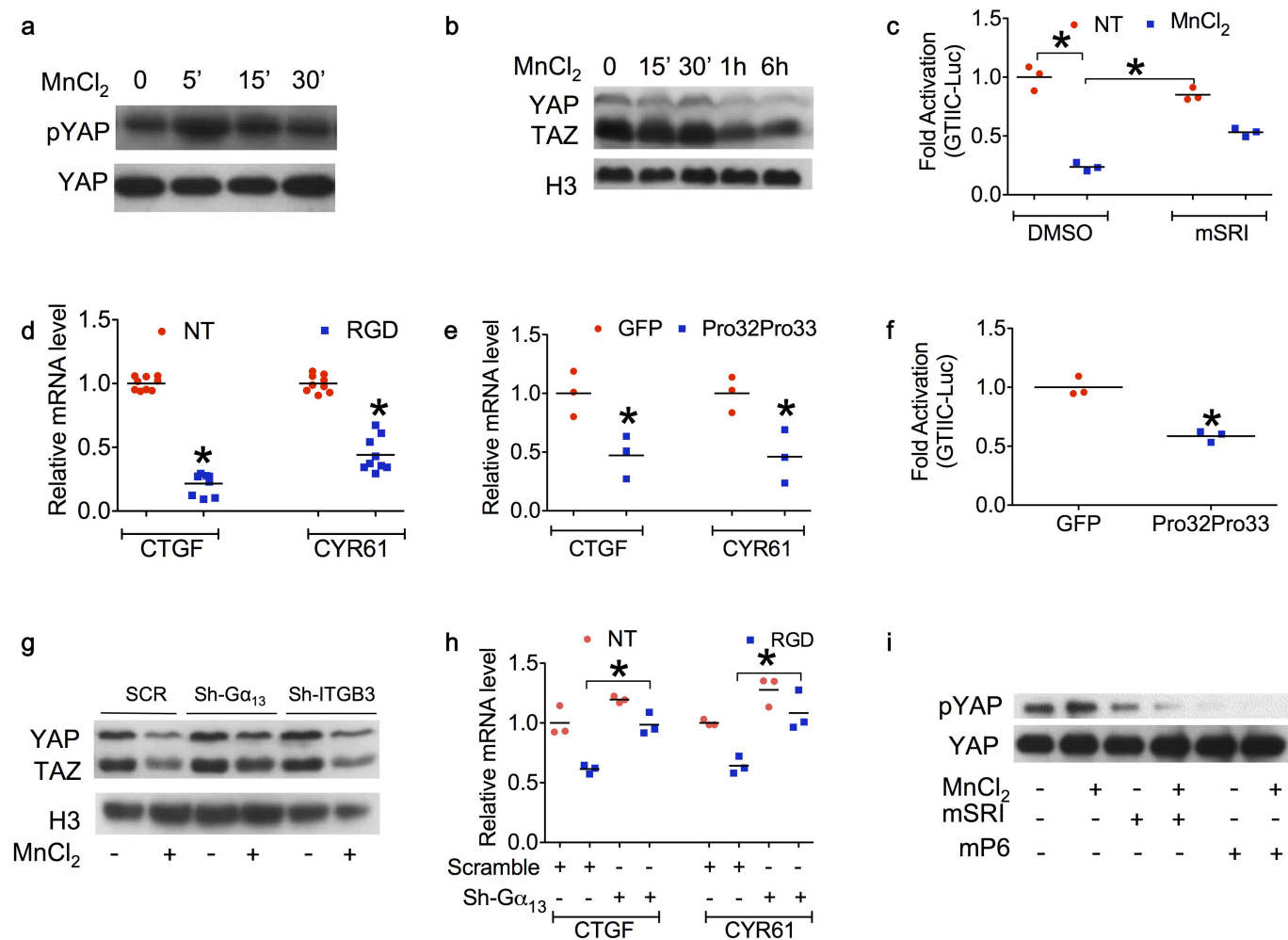
Data availability. The RNA-seq data that support the findings of this study have been deposited in BioSamples database (<https://www.ebi.ac.uk/biosamples/>) under accession number SAMN04565728. All other data are available from the corresponding authors upon reasonable request.

29. Wang, N. *et al.* Shear stress regulation of Krüppel-like factor 2 expression is flow pattern-specific. *Biochem. Biophys. Res. Commun.* **341**, 1244–1251 (2006).
30. Miao, H. *et al.* Effects of flow patterns on the localization and expression of VE-cadherin at vascular endothelial cell junctions: *in vivo* and *in vitro* investigations. *J. Vasc. Res.* **42**, 77–89 (2005).
31. Sun, X. *et al.* Activation of integrin $\alpha 5$ mediated by flow requires its translocation to membrane lipid rafts in vascular endothelial cells. *Proc. Natl Acad. Sci. USA* **113**, 769–774 (2016).
32. Zhao, B. *et al.* Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. *Genes Dev.* **21**, 2747–2761 (2007).
33. Varelak, X. *et al.* TAZ controls Smad nucleocytoplasmic shuttling and regulates human embryonic stem-cell self-renewal. *Nature Cell Biol.* **10**, 837–848 (2008).
34. Takagi, J., Petre, B. M., Walz, T. & Springer, T. A. Global conformational rearrangements in integrin extracellular domains in outside-in and inside-out signaling. *Cell* **110**, 599–611 (2002).
35. Luo, J. *et al.* A protocol for rapid generation of recombinant adenoviruses using the AdEasy system. *Nature Protocols* **2**, 1236–1247 (2007).
36. Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J. & Mateo, J. L. CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction Tool. *PLoS ONE* **10**, e0124633 (2015).
37. Varadi, K. *et al.* Novel random peptide libraries displayed on AAV serotype 9 for selection of endothelial cell-directed gene transfer vectors. *Gene Ther.* **19**, 800–809 (2012).
38. Grieger, J. C., Choi, V. W. & Samulski, R. J. Production and characterization of adeno-associated viral vectors. *Nature Protocols* **1**, 1412–1428 (2006).



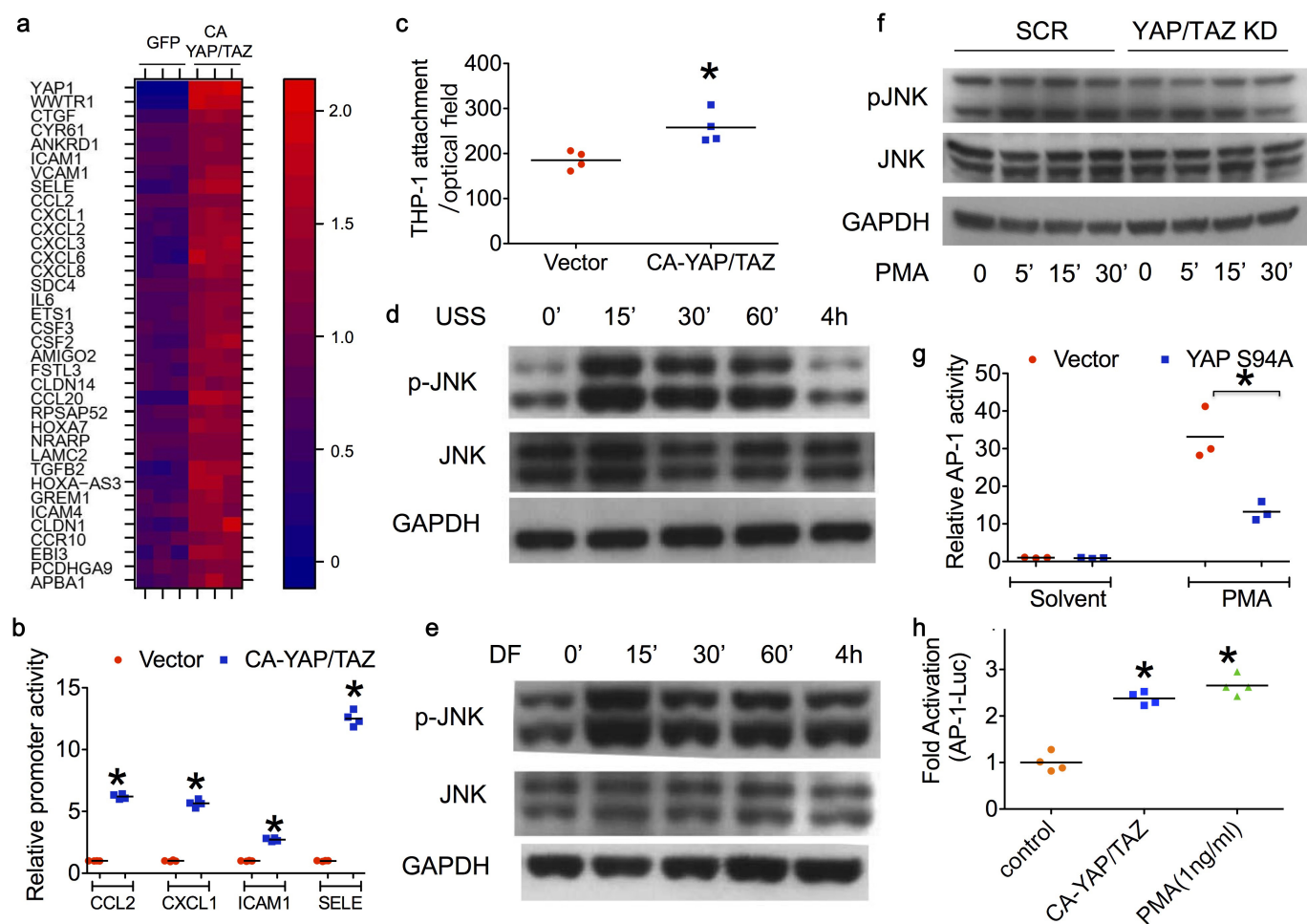
Extended Data Figure 1 | USS and disturbed flow oppositely regulate YAP/TAZ activity. **a**, Immunoblotting showing USS induces YAP phosphorylation in human aortic ECs. **b**, Summarized data for USS-induced YAP nuclear exportation ($n = 5$; * $P < 0.05$ by two-tailed unpaired t -test). **c**, TAZ is decreased in nuclear fractions and increased in cytoplasmic fractions in HUVECs exposed to USS for 6 h. TAZ expression was detected by immunoblotting after cell fractionation. **d**, Disturbed flow suppresses YAP phosphorylation in human aortic ECs.

e, Immunoblotting showing disturbed flow increases CTGF expression in HUVECs. All immunoblotting experiments were repeated three times and the representative results are shown. **f**, **g**, YAP/TAZ knockdown attenuates gene expression of disturbed-flow-induced (**f**) CTGF and (**g**) CYR61 ($n = 3$; * $P < 0.05$ by two-tailed unpaired t -test). NS, not significant. **h**, Summarized data for *en face* staining of relative nuclear YAP level in mouse aorta ($n_{TA} = 6$, $n_{AA, inner} = 3$, $n_{AA, outer} = 3$; * $P < 0.05$ by two-tailed unpaired t -test).



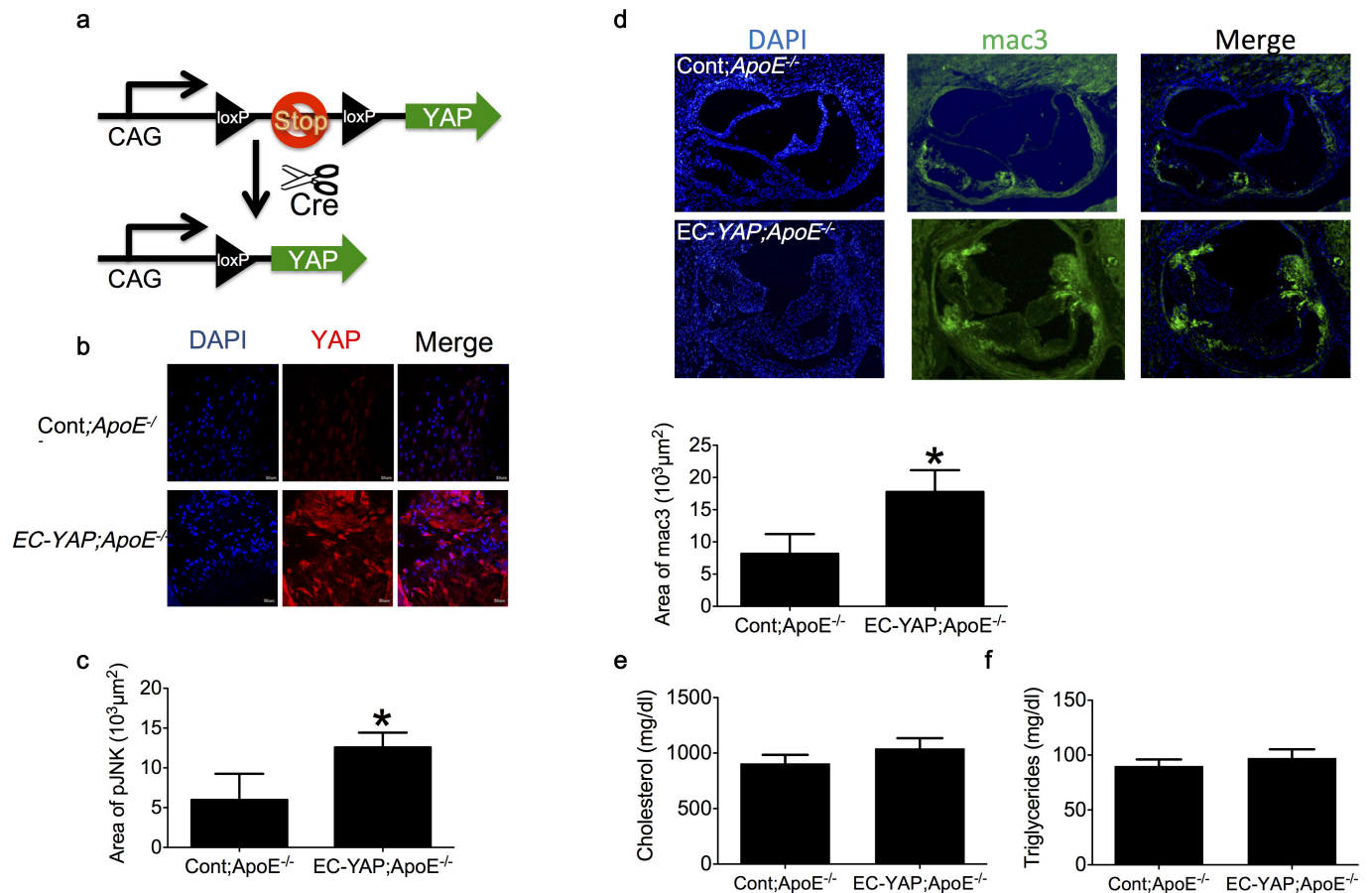
Extended Data Figure 2 | USS inhibits YAP/TAZ through integrin- $G\alpha_{13}$ -RhoA pathway. **a**, MnCl₂ (0.5 mM) promotes YAP phosphorylation shown by immunoblotting. **b**, MnCl₂ reduces nuclear YAP/TAZ levels in HUVECs. **c**, $G\alpha_{13}$ inhibiting peptide mSRI reverses MnCl₂-induced YAP/TAZ reporter (8×GTIIC-luc) gene activity ($n = 3$; $*P < 0.05$ by two-tailed unpaired t -test). **d**, RGD-containing peptide GRGDSP downregulates YAP/TAZ downstream target gene expression ($n = 3$; $*P < 0.05$ by two-tailed unpaired t -test). **e**, **f**, Pro32pro33 mutation in integrin β_3 inhibits YAP/TAZ transactivation in HUVECs, as verified

by suppressed (**e**) expression of YAP/TAZ target genes and (**f**) YAP/TAZ reporter gene activity ($n = 3$; $*P < 0.05$ by two-tailed unpaired t -test). **g**, $G\alpha_{13}$ or integrin β_3 knockdown reverses MnCl₂-induced YAP/TAZ nuclear exportation in HUVECs. **h**, $G\alpha_{13}$ knockdown reverses RGD-containing peptide-mediated CTGF and CYR61 suppression in HUVECs ($n = 3$; $*P < 0.05$ by two-tailed unpaired t -test). **i**, $G\alpha_{13}$ inhibiting peptide mSRI and mP6 reverse MnCl₂-induced (5 min) pYAP but not total YAP expression in HUVECs. The experiments were repeated at least three times and the representative results are shown.



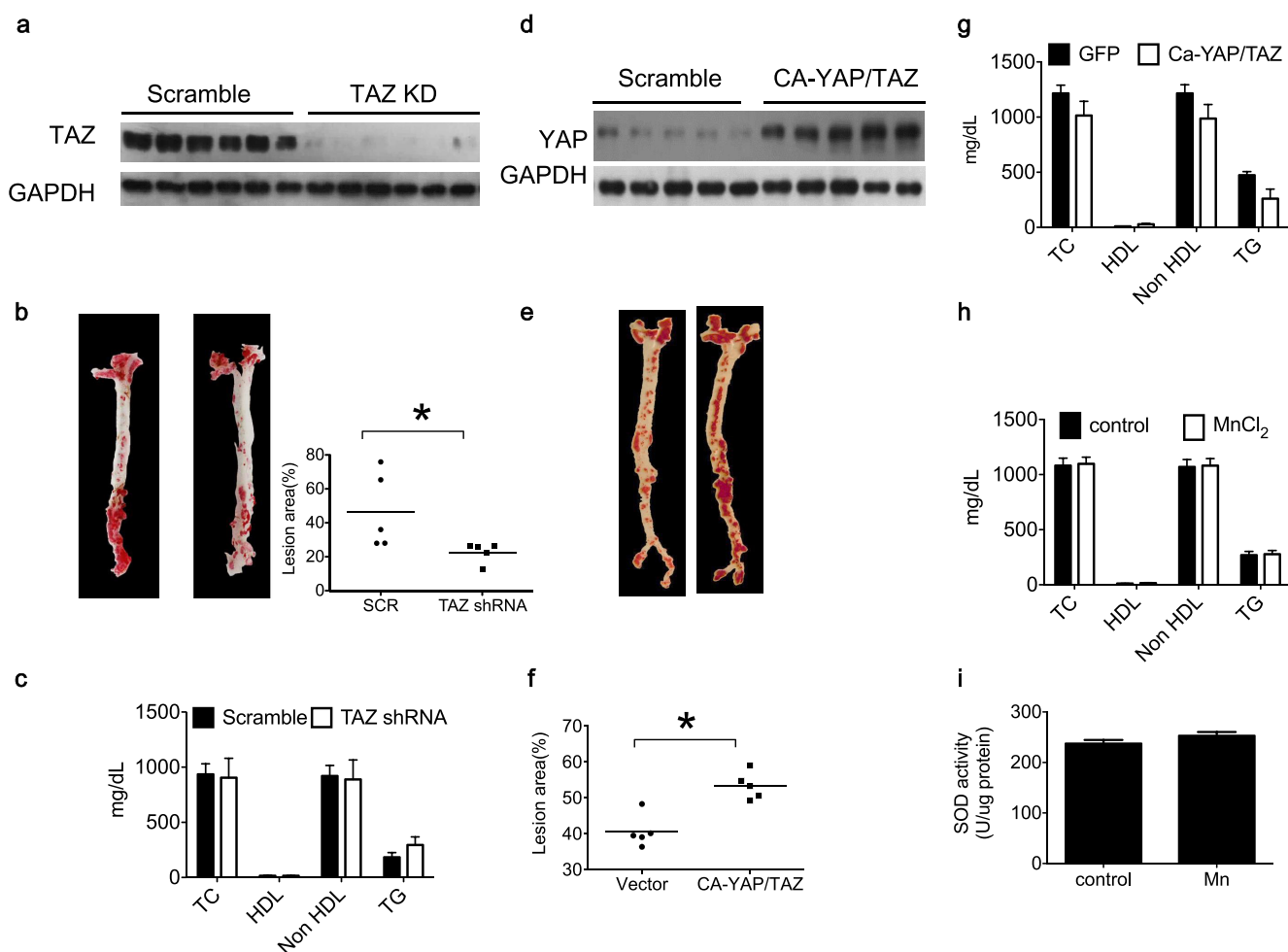
Extended Data Figure 3 | YAP/TAZ activation increases JNK activity. **a**, Heatmap for mRNA sequencing results showing CA-YAP/TAZ promotes expression of pro-inflammatory genes. **b**, CA-YAP/TAZ increases the promoter activity of adhesion molecules in HUVECs ($n = 4$; $*P < 0.05$ by two-tailed unpaired t -test). **c**, Summarized data for CA-YAP/TAZ overexpression increases monocyte attachment to HUVECs ($n = 4$; $*P < 0.05$ by two-tailed unpaired t -test). **d**, **e**, Immunoblotting showing JNK phosphorylation in HUVECs exposed to **(d)** USS or

(e) disturbed flow for different durations. Experiments were repeated three times and the representative results are shown. **f**, YAP/TAZ knockdown suppresses basal and PMA-induced JNK phosphorylation in HUVECs. **g**, Overexpression of dominant-negative YAP (YAP S94A) inhibits PMA-induced AP-1 reporter gene activity ($n = 3$; $*P < 0.05$ by two-tailed unpaired t -test). **h**, CA-YAP/TAZ increases AP-1 reporter gene activity in HUVECs ($n = 4$; $*P < 0.05$ by two-tailed unpaired t -test), and PMA was used as positive control for monitoring AP-1 activity.



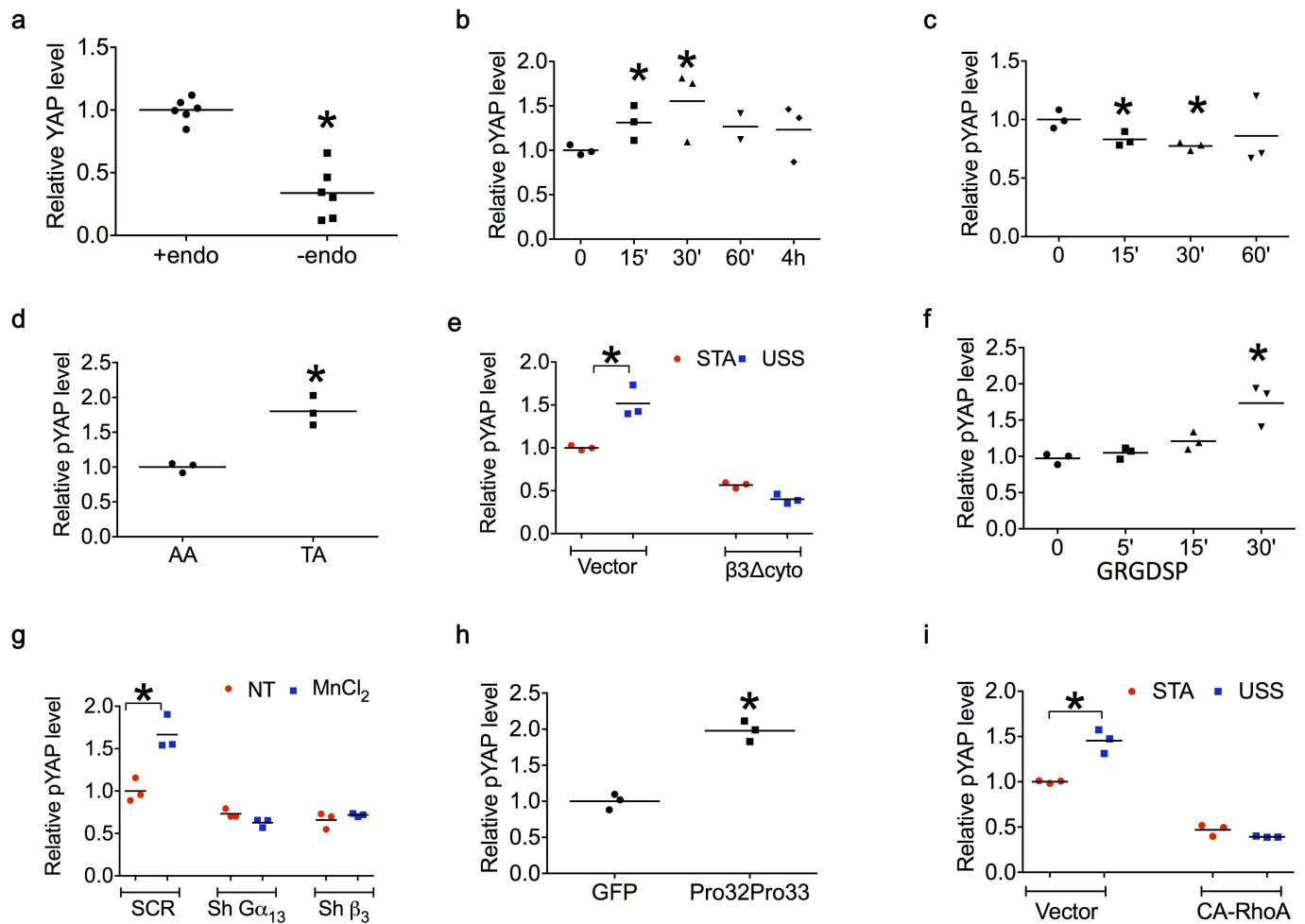
Extended Data Figure 4 | EC-specific overexpression of YAP accelerates plaque formation. **a**, The generation of Cre-mediated EC-specific YAP overexpression transgenic mice. **b**, *En face* staining showing increased YAP expression in endothelial cells of the *Tie2^{Cre/+};Yap-COE^{tg/+};ApoE^{-/-}* ($n = 10$). **c**, Summarized data for EC-specific YAP overexpression-increased JNK phosphorylation ($n = 10$; * $P < 0.05$ by two-tailed unpaired

t-test). **d**, EC-specific YAP overexpression increases macrophage content in the atherosclerotic plaques from aortic root ($n = 10$; * $P < 0.05$ by two-tailed unpaired *t*-test). **e**, **f**, EC-specific YAP overexpression does not affect serum levels of (e) cholesterol or (f) triglycerides ($n = 10$; * $P < 0.05$ by two-tailed unpaired *t*-test).



Extended Data Figure 5 | Inhibiting TAZ activity by shRNA or MnCl₂ administration delays atherogenesis and is independent of lipid metabolism, while activating YAP/TAZ by AAV-mediated CA-YAP/TAZ overexpression accelerates atherosclerotic plaque formation. **a**, Immunoblotting showing adenovirus-mediated TAZ shRNA suppressed TAZ expression level. **b**, TAZ knockdown delayed Western-diet-induced plaque formation in *ApoE*^{-/-} mice, $n = 5$; $*P < 0.05$ by two-tailed unpaired t -test. **c**, TAZ knockdown-suppressed plaque formation in *ApoE*^{-/-} mice is not due to change in lipid profile. Data are expressed as mean \pm s.e.m.,

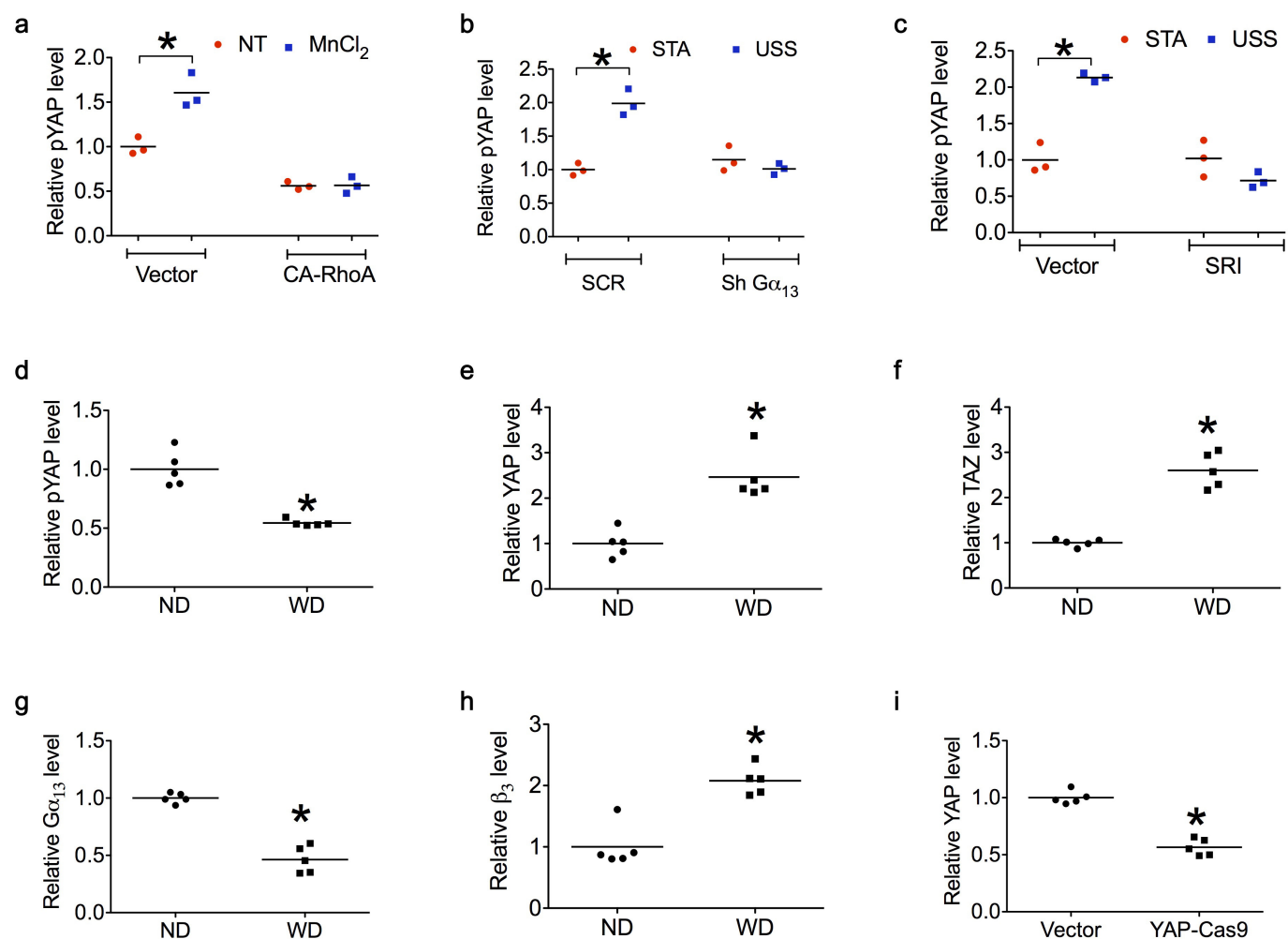
$n = 5$; $*P < 0.05$ by two-tailed unpaired t -test. **d**, Immunoblotting showing increased YAP expression in mice injected with AAV expressing CA-YAP/TAZ. **e**, **f**, Oil Red O staining (**e**) and summarized data (**f**) for CA-YAP/TAZ-induced exacerbation of plaque formation; $n = 5$, $*P < 0.05$ by two-tailed unpaired t -test. **g**, AAV-mediated CA-YAP/TAZ overexpression does not affect lipid profile in *ApoE*^{-/-} mice. **h**, **i**, Oral administration of MnCl₂ does not affect (**h**) lipid profile or (**i**) SOD activity in liver. Data are expressed as mean \pm s.e.m., $n = 5$; $*P < 0.05$ by two-tailed unpaired t -test.



Extended Data Figure 6 | Summary of western blotting data.

a, Endothelium removal reduces YAP level in mouse aorta. **b**, USS increases YAP phosphorylation. **c**, Disturbed flow reduces YAP phosphorylation. **d**, Thoracic aorta expresses higher levels of pYAP than aortic arch. **e**, Overexpression loss-of-function mutation of integrin β_3 ($\beta_3\Delta\text{cyto}$) suppresses USS-induced pYAP. **f**, RGD-containing peptide

GRGDSP induces pYAP. **g**, $\text{G}\alpha_{13}$ or integrin β_3 knockdown reverses MnCl_2 -induced pYAP. **h**, Integrin gain-of-function mutation Pro32Pro33 increases pYAP. **i**, Constitutively activated RhoA (CA-RhoA) reverses USS-induced pYAP. Data: $n = 6$ for **a** and $n = 3$ for other figures; $*P < 0.05$ by two-tailed unpaired t -test.



Extended Data Figure 7 | Summary of western blotting data.

a, CA-RhoA reverses MnCl₂-induced pYAP. **b**, Gα₁₃ knockdown reverses USS-induced pYAP. **c**, Gα₁₃ inhibitor SRI reverses USS-induced pYAP. **d–h**, Immunoblotting detection of **(d)** pYAP, **(e)** YAP, **(f)** TAZ,

(g) Gα₁₃ and **(h)** integrin β₃ levels. **i**, YAP knockdown by the CRISPR-Cas9 *in vivo* genome editing system. Data: $n = 3$ for **a–c**, $n = 5$ for **d–i**; * $P < 0.05$ by two-tailed unpaired *t*-test.

Extended Data Table 1 | Drugs and concentrations used for the YAP/TAZ inhibition test

Drugs	Stock concentration	Working concentration	Company
Adenosine	10 mM	10 μ M	Sigma
Apelin	1 mM	100 nM	Sigma
Exendin 4	100 μ M	10 nM	Sigma
Nicotinic acid (VB3 niacin)	500 mM	3 to 5 mM	Sigma
ApoA1	1 mg/mL	10 μ g/mL	Sigma
Rosuvastatin	10 mM	10 μ M	Cayman
Simvastatin	100 mM	1 μ M	Cayman
Atorvastatin	10 mM	1 μ M	Cayman

NLRC3 is an inhibitory sensor of PI3K–mTOR pathways in cancer

Rajendra Karki^{1*}, Si Ming Man^{1*}, R. K. Subbarao Malireddi¹, Sannula Kesavardhana¹, Qifan Zhu^{1,2}, Amanda R. Burton¹, Bhesh Raj Sharma¹, Xiaopeng Qi¹, Stephane Pelletier^{1,3}, Peter Vogel⁴, Philip Rosenstiel⁵ & Thirumala-Devi Kanneganti¹

NLRs (nucleotide-binding domain and leucine-rich repeats) belong to a large family of cytoplasmic sensors that regulate an extraordinarily diverse range of biological functions. One of these functions is to contribute to immunity against infectious diseases, but dysregulation of their functional activity leads to the development of inflammatory and autoimmune diseases¹. Cytoplasmic innate immune sensors, including NLRs, are central regulators of intestinal homeostasis^{2–9}. NLRC3 (also known as CLR16.2 or NOD3) is a poorly characterized member of the NLR family and was identified in a genomic screen for genes encoding proteins bearing leucine-rich repeats (LRRs) and nucleotide-binding domains^{10,11}. Expression of NLRC3 is drastically reduced in the tumour tissue of patients with colorectal cancer compared to healthy tissues¹², highlighting an undefined potential function for this sensor in the development of cancer. Here we show that mice lacking NLRC3 are hyper-susceptible to colitis and colorectal tumorigenesis. The effect of NLRC3 is most dominant in enterocytes, in which it suppresses activation of the mTOR signalling pathways and inhibits cellular proliferation and stem-cell-derived organoid formation. NLRC3 associates with PI3Ks and blocks activation of the PI3K-dependent kinase AKT following binding of growth factor receptors or Toll-like receptor 4. These findings reveal a key role for NLRC3 as an inhibitor of the mTOR pathways, mediating protection against colorectal cancer.

Previous studies have shown that NLRC3 functions as a negative regulator of signalling pathways activated by Toll-like receptors (TLRs) and the DNA sensor STING in response to pathogen-associated molecular patterns or to virus infection^{13,14}. The physiological role of NLRC3 has, however, remained largely unknown. Using an established mouse model of colitis-associated colorectal tumorigenesis, we investigated the role of NLRC3 in colorectal cancer. To do this, we injected mice intraperitoneally with azoxymethane, followed by three rounds of dextran sulfate sodium (DSS) treatment (Extended Data Fig. 1a). All time points referred to hereafter indicate the number of days after injection of azoxymethane. The number of tumours was quantified at day 80. Qualitative reverse-transcriptase PCR analysis revealed a reduction in the expression of the gene encoding NLRC3 in tumour tissue compared with non-tumour-associated tissue in the colon of wild-type mice 80 days after injection of azoxymethane (day 80; Extended Data Fig. 1b).

We injected cohorts of co-housed wild-type and *Nlrc3*^{−/−} mice with azoxymethane, followed by three rounds of DSS treatment, and examined the prevalence of tumours in the colon of these mice at day 80 (Extended Data Fig. 1a, c). We found that *Nlrc3*^{−/−} mice lost more body weight after the first two rounds of DSS treatment compared to wild-type mice and developed significantly more tumours (Fig. 1a–d). Histological hallmarks associated with thickening of the colon, inflammation, ulceration, hyperplasia and the extent or severity of damage were more frequently identified in the middle and distal

colon and the rectum of *Nlrc3*^{−/−} mice compared to the corresponding regions in wild-type mice (Fig. 1e, f and Extended Data Fig. 1d).

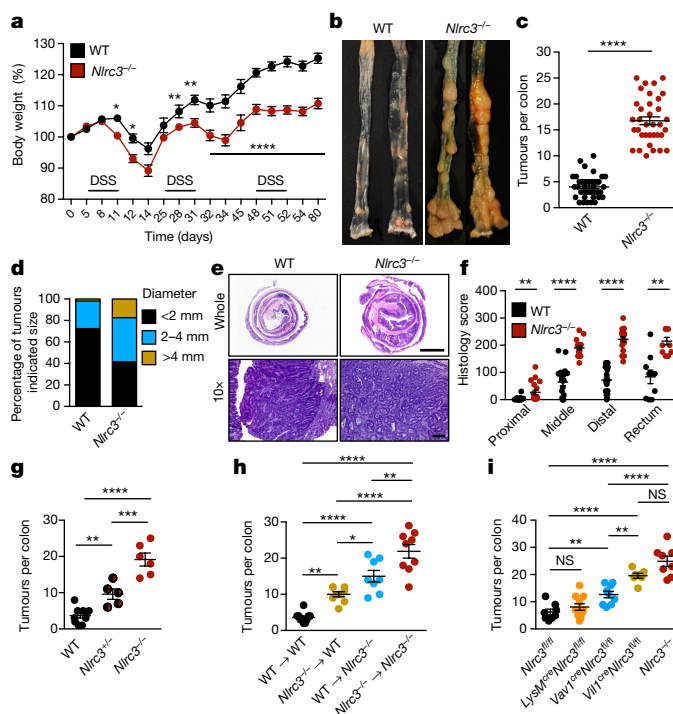


Figure 1 | NLRC3 prevents colorectal tumorigenesis. **a**, Percentage change in body weight during azoxymethane and DSS treatment; day 0 is the time of azoxymethane injection. **b**, Representative images of colon tumours in mice 80 days after injection of azoxymethane. **c**, **d**, Number (**c**) and size (**d**) of colon tumours in wild-type (WT; $n = 37$) and *Nlrc3*^{−/−} ($n = 35$) mice. **e**, **f**, Haematoxylin and eosin staining (**e**) and histological scores (**f**) of tumours in colon as shown in **b**. Scale bars, 2,500 μm (whole colon, top), 200 μm (magnified, bottom). **g**, Number of colon tumours in littermate wild-type ($n = 10$), *Nlrc3*^{+/−} ($n = 5$) and *Nlrc3*^{−/−} ($n = 6$) mice. **h**, Number of colon tumours in bone-marrow chimaera mice treated as in **b**. Wild-type mice transplanted with wild-type bone marrow ($n = 10$); *Nlrc3*^{−/−} mice transplanted with wild-type bone marrow ($n = 9$); wild-type mice transplanted with *Nlrc3*^{−/−} bone marrow ($n = 8$); and *Nlrc3*^{−/−} transplanted with *Nlrc3*^{−/−} bone marrow ($n = 9$). **i**, Number of colon tumours in littermate *Nlrc3*^{fl/fl} ($n = 8$), *LysM*^{cre}*Nlrc3*^{fl/fl} ($n = 11$), *Vav1*^{cre}*Nlrc3*^{fl/fl} ($n = 9$), *Vil1*^{cre}*Nlrc3*^{fl/fl} ($n = 7$) and *Nlrc3*^{−/−} ($n = 8$) mice treated as in **a**. Each symbol represents an individual mouse (**c**, **f**–**i**). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$; NS, not statistically significant by one-way ANOVA (**a**, **g**–**i**) or two-tailed *t*-test (**c**, **f**). Data are from three (**a**–**f**) or two independent experiments (**g**–**i**) and are presented as mean \pm s.e.m. in **a**, **c**, **f**–**i**.

¹Department of Immunology, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ²Integrated Biomedical Sciences Program, University of Tennessee Health Science Center, Memphis, Tennessee 38163, USA. ³Embryonic Stem Cell Laboratory, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ⁴Animal Resources Center and the Veterinary Pathology Core, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ⁵Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, D-24105 Kiel, Germany.

*These authors contributed equally to this work.

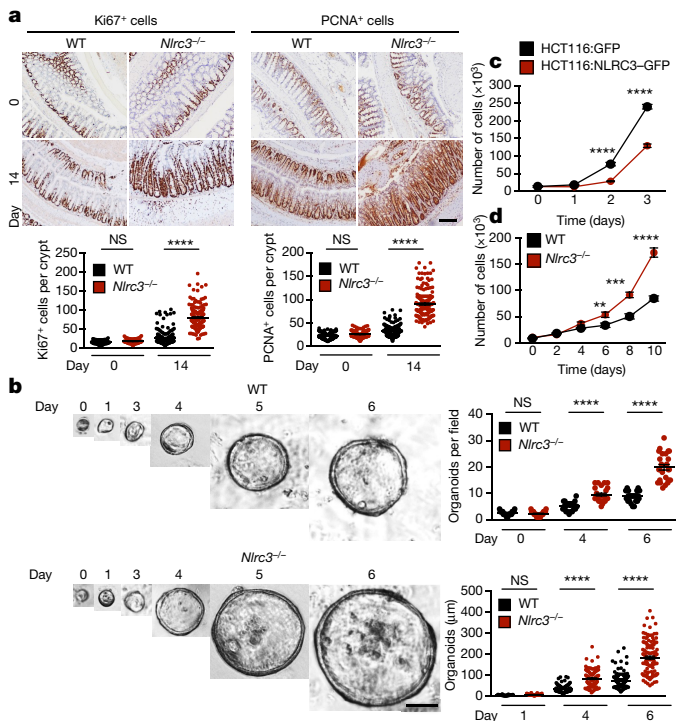


Figure 2 | NLRC3 suppresses overt proliferation. **a**, Images and quantification of the number of Ki67⁺ (left) and PCNA⁺ (right) cells in each crypt of wild-type (day 0, $n = 5$; day 14, $n = 8$) and *Nlr3*^{-/-} (day 0, $n = 5$; day 14, $n = 8$) mice. **b**, Images (left) and quantifications (right) of the number (top) and size (bottom) of mouse intestinal organoids. **c**, Proliferation of the HCT116 cell line. **d**, Proliferation of primary mouse fibroblasts. At least 25 crypts were counted in each animal (**a**). Scale bars, 200 μm (**a**) or 50 μm (**b**). Each symbol represents one crypt (**a**) or one organoid (**b**). ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$; NS, not statistically significant by two-tailed t -test. Data are from one experiment representative of two (**a**, **b**) or three independent experiments (**c**, **d**) and are presented as mean \pm s.e.m. in **a–d**.

All *Nlr3*^{-/-} mice suffered high-grade dysplasia, whereas wild-type mice suffered low-grade dysplasia (Extended Data Fig. 1e). We found that 63% of the *Nlr3*^{-/-} mice were positive for adenocarcinoma in the colon, compared to 0% of the wild-type controls at day 80 (Extended Data Fig. 1f). Although NLRC3 showed a gene-dose-dependent response to azoxymethane and DSS (Fig. 1g), it does not appear to have an effect on the normal mouse intestine or colon (Extended Data Fig. 1g).

Nlr3^{-/-} mice lost significantly more body weight and suffered more severe shortening of, and damage to, the colon after only a single round of DSS treatment compared to their wild-type counterparts (Extended Data Fig. 2a–c). Certain members of the NLR protein family can form inflammasomes, driving maturation of IL-18, a cytokine important for mediating protection against colitis-associated tumorigenesis^{2–4,15}. We did not observe differential production of IL-18 in wild-type and *Nlr3*^{-/-} mice at day 14 or at day 80 (Extended Data Fig. 2d). Instead, production of the other inflammasome-associated cytokine IL-1 β and inflammasome-independent cytokines IL-6, TNF and GCSF and the chemokines KC (also known as CXCL1), MCP-1 (also known as CCL2) and MIP-1 α (also known as CCL3) was elevated in colon tissue of *Nlr3*^{-/-} mice compared to wild-type mice at day 14 (Extended Data Fig. 2d–g). We further confirmed these results and also found increased levels of circulating IL-6, GCSF, KC and MIP-1 α in the sera of *Nlr3*^{-/-} mice compared to wild-type mice at day 14 (Extended Data Fig. 2h). The expression of IL-17 and IL-22 was also elevated in the colon tissue of *Nlr3*^{-/-} mice compared to wild-type mice at day 14, whereas the expression of IL-23, IFN β and IFN γ remained unchanged (Extended Data Fig. 3a). Consistent with the observation that *Nlr3*^{-/-} mice had

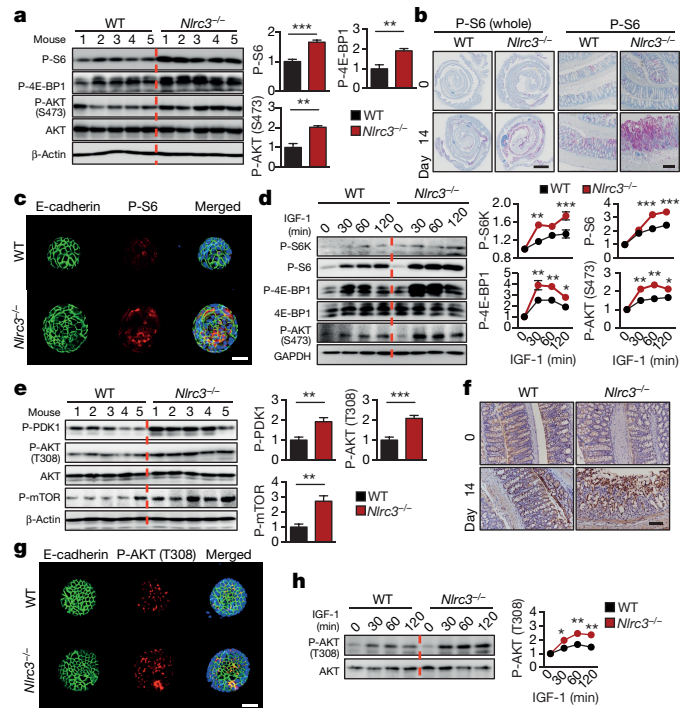


Figure 3 | NLRC3 controls mTOR signalling pathways. **a**, Immunoblot of mouse colon tissue and densitometric quantification at day 14. P- indicates phosphorylated protein. **b**, Immunohistochemical staining of mouse colon tissue. **c**, Immunofluorescence staining of mouse intestinal organoids after 7 days of culture. **d**, Immunoblot of organoids treated with IGF-1 and densitometric quantification. **e**, Immunoblot of colon tissue of mice and densitometric quantification at day 14. **f**, Immunohistochemical staining of colon tissue of mice for AKT that has been phosphorylated at Ser473. **g**, Immunofluorescence staining of mouse intestinal organoids after 7 days of culture. **h**, Immunoblot of organoids treated with IGF-1 and densitometric quantification. Scale bars, 2,500 μm (**b**, whole colon, left), 200 μm (**b**, magnified, right), 50 μm (**c**, **f**, **g**). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; by two-tailed t -test (**a** and **e**) or one-way ANOVA (**d**, **h**). Data are from one experiment representative of two (**a–c**, **e–g**) or three independent experiments (**d**, **h**) and are presented as mean \pm s.e.m. in **a**, **d**, **e** and **h**. For gel source data, see Supplementary Fig. 1.

elevated levels of many pro-inflammatory mediators at day 14, we observed increased levels of I κ B α and STAT3 phosphorylation in the colon tissue of *Nlr3*^{-/-} mice compared to wild-type mice (Extended Data Fig. 3b). However, differential phosphorylation of ERK was not observed (Extended Data Fig. 3b). Global increases in the production of inflammatory mediators and the activation of immune signalling pathways reflected the hyper-susceptibility of *Nlr3*^{-/-} mice to colitis.

Using flow cytometry, we profiled the immune cell populations in the colons of untreated wild-type and *Nlr3*^{-/-} mice and wild-type and *Nlr3*^{-/-} mice at days 8 and 14. We observed an increased number of macrophages, neutrophils and natural killer cells in the colons of *Nlr3*^{-/-} mice compared to wild-type mice 14 days after azoxymethane and DSS treatment (Extended Data Fig. 3c, d), which is consistent with the increased levels of inflammation observed at this time point. However, we did not observe differences in the relative number of macrophages, CD11b⁺CD11c⁺ cells, neutrophils, B cells, CD4⁺ T cells, CD8⁺ T cells and natural killer cells between untreated wild-type and *Nlr3*^{-/-} mice or mice at day 8 (Extended Data Fig. 3d). NLRC3 has been implicated in the regulation of T-cell activation¹¹; however, we did not observe a difference in the levels of IFN γ ⁺ or TNF⁺ CD4⁺ T cells when wild-type and *Nlr3*^{-/-} splenocytes were stimulated with CD3 and CD28 in the presence of IL-2 (Extended Data Fig. 3e).

We performed bone-marrow chimaera studies to identify the contribution of NLRC3 in haematopoietic cells versus radioresistant stromal

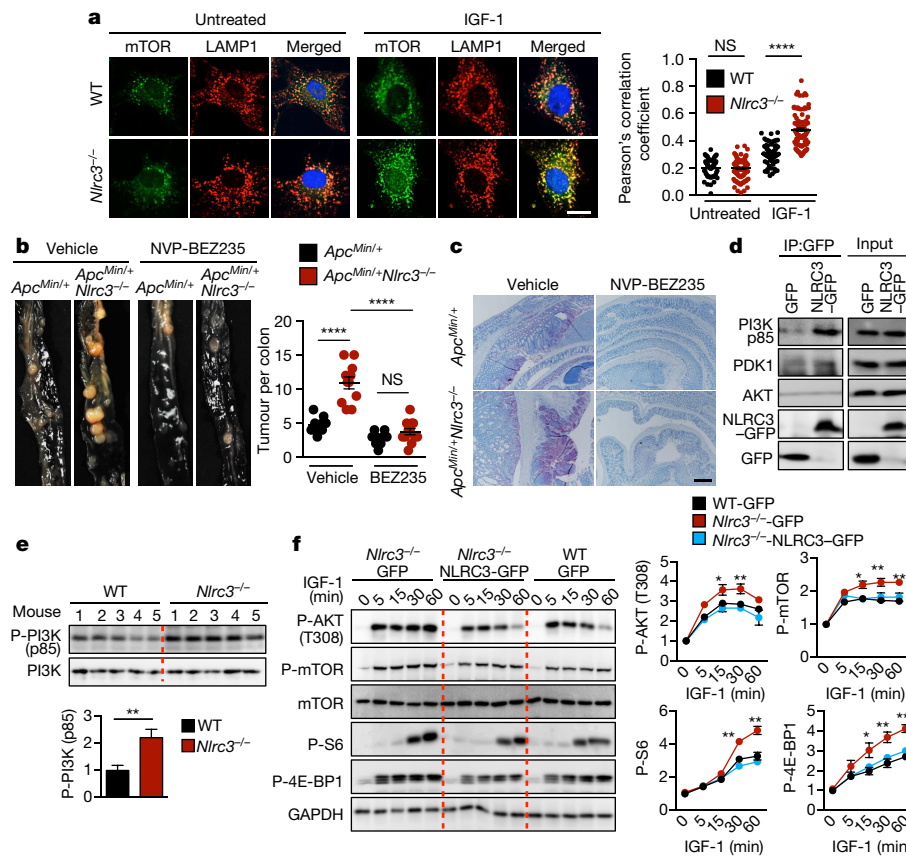


Figure 4 | NLRC3 regulates upstream signalling molecules within the PI3K-AKT-mTOR pathway. **a**, Immunofluorescence staining of primary fibroblasts and frequency of co-localization between mTOR and LAMP1 ($n \geq 150$). **b**, Images and quantification of colon tumours in littermate *Apc*^{Min/+} and *Apc*^{Min/+} *Nlrc3*^{-/-} mice 40 days after treatment with vehicle or NVP-BEZ235. **c**, Immunohistochemical staining for phosphorylated S6 of colon tissue from mice treated as in **b**. **d**, Immunoprecipitation of the GFP tag in 239T cells transfected with a plasmid encoding GFP alone or NLRC3-GFP. **e**, Immunoblot of mouse colon tissue and densitometric

quantification at day 14. **f**, Immunoblot of primary mouse fibroblasts transduced with a retroviral vector encoding GFP or NLRC3-GFP, with or without stimulation with IGF-1, and densitometric quantification. Scale bar, 10 μ m (**a**), 200 μ m (**c**). * $P < 0.05$; ** $P < 0.01$; **** $P < 0.0001$; NS, no statistical significance by two tailed *t*-test (**e**) or one-way ANOVA (**a**, **b**, **f**). Data are from one experiment representative of two (**a**–**c**, **e**) or three (**d** and **f**) independent experiments (mean \pm s.e.m. in **a**, **b**, **e**, **f**). For gel source data, see Supplementary Fig. 1.

cells during colitis-associated tumorigenesis. As we expected, *Nlrc3*^{-/-} mice that received *Nlrc3*^{-/-} bone marrow were more susceptible to tumorigenesis than wild-type mice that received wild-type bone marrow (Fig. 1h and Extended Data Fig. 4a). However, *Nlrc3*^{-/-} mice that received wild-type bone marrow had a significantly increased tumour burden compared with wild-type mice that received wild-type bone marrow. In addition, wild-type mice that received *Nlrc3*^{-/-} bone marrow had a significantly increased tumour burden compared to wild-type mice that received wild-type bone marrow (Fig. 1h and Extended Data Fig. 4a). We further confirmed our findings and generated mice that lacked NLRC3 in haematopoietic cells specifically (*Vav1*^{cre}*Nlrc3*^{fl/fl}), cells of the myeloid lineage (*LysM*^{cre}*Nlrc3*^{fl/fl}) and intestinal epithelial cells (*Vil1*^{cre}*Nlrc3*^{fl/fl}). Mice lacking NLRC3 in intestinal epithelial cells developed the highest number of tumours, followed by mice lacking NLRC3 in haematopoietic cells (Fig. 1i and Extended Data Fig. 4b). Mice lacking NLRC3 in cells of the myeloid lineage had a similar number of tumours to wild-type mice (Fig. 1i and Extended Data Fig. 4b). These data support the observation that the oncogenic inhibitory effect of NLRC3 is more dominant in intestinal epithelial cells and more subtle in haematopoietic cells.

A closer examination of the intestinal epithelial cells of *Nlrc3*^{-/-} mice revealed a significant increase in numbers of both Ki67⁺ and PCNA⁺ (both proteins that are associated with cellular proliferation) cells per intestinal crypt compared to wild-type mice at day 14 (Fig. 2a). Additionally, colonic epithelial stem cells collected from *Nlrc3*^{-/-} mice more readily developed into organoids in *ex vivo* culture compared to

those collected from wild-type mice (Fig. 2b). The average diameters of the organoids derived from *Nlrc3*^{-/-} mice were significantly increased compared to the average diameters of the organoids from wild-type mice (Fig. 2b). Expression of the stem-cell marker *Lgr5* in the colon tissue was similar in wild-type and *Nlrc3*^{-/-} mice (Extended Data Fig. 4c), suggesting that the differences in the number and size of intestinal organoids was due to differential colony-forming capacity rather than differences in the numbers of starting intestinal stem cells. To investigate the effect of NLRC3 on cell proliferation more directly, we overexpressed NLRC3 in the human colon cell line HCT116 and found that these cells exhibited reduced levels of proliferation compared to cells expressing a control (GFP) protein (Fig. 2c). Furthermore, primary *Nlrc3*^{-/-} fibroblasts proliferated more rapidly than wild-type fibroblasts (Fig. 2d).

Cellular proliferation can also be achieved when growth factors, nutrients and cellular energy activate metabolic pathways via the kinase mTOR¹⁶. We found increased phosphorylation levels of S6 kinase, 4E-BP1 and AKT at Ser473, the downstream targets of mTOR, in the colon tissue of *Nlrc3*^{-/-} mice compared to wild-type mice at day 14 (Fig. 3a, b). Increased phosphorylation of these mTOR targets was also observed in *Nlrc3*^{-/-} organoids compared to wild-type organoids (Fig. 3c, d). However, we found no difference in the expression of genes involved in the Wnt signalling pathway, including *Wnt1*, *Ctnnb1*, *Lef1*, *Tcf4*, *Tcf7* and *Axin2* (Extended Data Fig. 4c). The extent of nuclear localization of β -catenin was also similar between wild-type and *Nlrc3*^{-/-} mice (Extended Data Fig. 4d).

Dysregulation of the mTOR signalling pathways in *Nlrc3*^{-/-} mice occurred very early, 8 days after injection of azoxymethane, whereas no difference in the production of inflammatory cytokines and mediators and phosphorylation of IκBα was observed at this time point (Extended Data Fig. 5, a–c). Moreover, the differential AKT–mTOR signalling observed at this time point was not caused by differences in the relative levels of immune cells recruited to the colon (Extended Data Fig. 3d). These data suggest that the dysregulated mTOR signalling observed at an earlier time point may lead to increased NF-κB signalling at a later time point.

Phosphorylation and activation of mTOR is driven by a number of upstream signalling proteins in the PI3K–AKT–mTOR pathway. Phosphorylation of AKT at the Thr308 site by the kinase PDK1 allows AKT to activate mTOR^{17–19}. To examine whether NLRC3 directly affects the apical molecules in the PI3K–AKT–mTOR pathway, we first investigated the phosphorylation status of AKT at the Thr308 site. We observed elevated phosphorylation of AKT at Thr308 in the colon tissue of *Nlrc3*^{-/-} mice compared to wild-type mice at day 14 (Fig. 3e). Increased levels of AKT phosphorylation in colon tissue were observed predominantly in epithelial cells and, to a lesser extent, in infiltrating cells (Fig. 3f). Similarly, increased phosphorylation of AKT at Thr308 was observed in *Nlrc3*^{-/-} organoids treated with IGF-1 compared to wild-type controls (Fig. 3g, h). We further observed increased activation of PDK1 in the colon tissue of *Nlrc3*^{-/-} mice treated with azoxymethane and DSS compared to wild-type mice (Fig. 3e). In addition, elevated degree of phosphorylation of AKT at Thr308 and of 4E-BP1 were observed in the colon of *Nlrc3*^{+/-} heterozygous mice compared to littermate wild-type mice at day 14, although this increase was smaller than that observed in littermate *Nlrc3*^{-/-} homozygous mice (Extended Data Fig. 6a). The gene-dose-dependent effect of NLRC3 on the suppression of the mTOR signalling pathways is reminiscent of the gene-dose-dependent effect of NLRC3 on the suppression of tumorigenesis (Fig. 1g).

Activated mTOR is phosphorylated and migrates to lysosomal and late endosomal membranes^{20,21}. We observed increased phosphorylation of mTOR in the colon tissue of *Nlrc3*^{-/-} mice (Fig. 3e). We also found an increased co-localization frequency between mTOR and LAMP1 puncta in IGF-1-treated primary *Nlrc3*^{-/-} fibroblasts compared to IGF-1-treated primary wild-type fibroblasts (Fig. 4a). Increased mTOR signalling was observed in *Nlrc3*^{-/-} fibroblasts or wild-type fibroblasts treated with short interfering RNAs (siRNAs) against *Nlrc3* compared to their corresponding controls (Extended Data Figs 6b–e, 7a–d).

We further investigated whether NLRC3 is able to restrict proliferation in a spontaneous mouse model of colon cancer. We crossed the mouse line containing a heterozygous mutation in the gene encoding adenomatous polyposis coli (*Apc*^{Min/+}) with *Nlrc3*^{-/-} mice and found that *Apc*^{Min/+}*Nlrc3*^{-/-} mice had a higher tumour burden than *Apc*^{Min/+} control mice (Extended Data Fig. 8a). Of the *Apc*^{Min/+}*Nlrc3*^{-/-} mice, 40% developed hyperplasia (compared to 0% in the littermate control group) and *Apc*^{Min/+}*Nlrc3*^{-/-} mice exhibited increased damage in the colon (Extended Data Fig. 8b, c). Notably, we observed increased number of Ki67⁺ proliferative cells and cells positive for phosphorylated S6 kinase in the colon of *Apc*^{Min/+}*Nlrc3*^{-/-} mice compared to *Apc*^{Min/+} mice (Extended Data Fig. 8c). Moreover, the capacity of *Apc*^{Min/+}*Nlrc3*^{-/-} intestinal stem cells to proliferate into organoids was greater than that of *Apc*^{Min/+} intestinal stem cells (Extended Data Fig. 8d). Treatment with the PI3K inhibitor LY294002 and the mTOR inhibitor rapamycin impaired the ability of intestinal stem cells to proliferate into organoids in both strains (Extended Data Fig. 8d). Treatment of the *Apc*^{Min/+} mice and *Apc*^{Min/+}*Nlrc3*^{-/-} mice with the PI3K–mTOR inhibitor NVP-BEZ235 reduced the tumour burden and phosphorylation of S6 kinase in the tumours and enterocytes of *Apc*^{Min/+}*Nlrc3*^{-/-} mice to a level observed in treated *Apc*^{Min/+} mice (Fig. 4b, c). Collectively, these data suggest that NLRC3 restricts cellular proliferation via the PI3K–mTOR axis during colon tumorigenesis.

Following generation of inositol phospholipids by activated PI3Ks, AKT is recruited to the cell membrane where it undergoes a conformational change and is phosphorylated at Thr308 by PDK1 (refs 17–19). Co-immunoprecipitation assays showed that NLRC3 weakly interacted with PDK1 and did not interact with AKT (Fig. 4d). Instead, we found that NLRC3 co-immunoprecipitated with p85 subunits of PI3K (Fig. 4d). In addition, we observed increased levels of interaction between the p85 and p110α subunits of PI3K in *Nlrc3*^{-/-} primary fibroblasts or mouse bone-marrow-derived macrophages (BMDMs) (Extended Data Fig. 9a, b). We also observed a higher level of phosphorylation and activation of p85 PI3K in the colon tissue of *Nlrc3*^{-/-} mice treated with azoxymethane and DSS compared to wild-type mice (Fig. 4e). These data provide evidence to suggest that NLRC3 disrupts an association between the PI3K p85 and p110α subunits and reduces the activity of PI3K p85 itself. Deletion of the CARD, NACHT or LRR domain of NLRC3 impaired the ability of NLRC3 to interact with either the p85 or the p110α subunit of PI3K (Extended Data Fig. 9c–g). Reconstitution of NLRC3 in *Nlrc3*^{-/-} fibroblasts reduced the degree of phosphorylation of AKT Thr308 and other downstream molecules to levels similar to those seen in wild-type fibroblasts upon stimulation with IGF-1 (Fig. 4f).

In addition to growth factor receptors, activation of TLR4 can engage the PI3K–AKT–mTOR pathway^{22–24}. We observed increased activation of the mTOR signalling pathways in lipopolysaccharide (LPS)-treated primary *Nlrc3*^{-/-} BMDMs compared to LPS-treated wild-type BMDMs (Extended Data Fig. 10a, b). We further confirmed our findings in an independently generated line of NLRC3-deficient mice that we term *Nlrc3*^{ld/ld} mice (NLRC3 large deletion, data not shown; see Methods). Collectively these findings identify NLRC3 as an inhibitory sensor of the PI3K–AKT–mTOR pathway, mediating protection against tumorigenesis in colorectal cancer (Extended Data Fig. 10c).

NLRC3 does not act solely to protect against cancer. A previous study has shown that expression of NLRC3 is downregulated in patients with the autoimmune disease Wegener's granulomatosis²⁵. Moreover, a loss-of-function mutation in the gene encoding the NLRC3-like protein in zebrafish results in systemic inflammation²⁶. These findings collectively provide evidence to support the cross-species functionality of NLRC3. Whether NLRC3 needs to be bound to a specific ligand or is engaged in the activation of its regulatory function in a ligand-independent manner remains to be explored. Understanding the precise functions of NLRC3 could open up new avenues in the treatment of infectious and autoinflammatory diseases and cancer.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 November 2015; accepted 2 November 2016.

Published online 12 December 2016.

1. Ting, J. P., Willingham, S. B. & Bergstralh, D. T. NLRs at the intersection of cell death and immunity. *Nat. Rev. Immunol.* **8**, 372–379 (2008).
2. Zaki, M. H. et al. The NLRP3 inflammasome protects against loss of epithelial integrity and mortality during experimental colitis. *Immunity* **32**, 379–391 (2010).
3. Zaki, M. H., Vogel, P., Body-Malapel, M., Lamkanfi, M. & Kanneganti, T. D. IL-18 production downstream of the Nlrp3 inflammasome confers protection against colorectal tumor formation. *J. Immunol.* **185**, 4912–4920 (2010).
4. Allen, I. C. et al. The NLRP3 inflammasome functions as a negative regulator of tumorigenesis during colitis-associated cancer. *J. Exp. Med.* **207**, 1045–1056 (2010).
5. Allen, I. C. et al. NLRP12 suppresses colon inflammation and tumorigenesis through the negative regulation of noncanonical NF-κB signaling. *Immunity* **36**, 742–754 (2012).
6. Elinav, E. et al. NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* **145**, 745–757 (2011).
7. Man, S. M. et al. Critical role for the DNA sensor AIM2 in stem cell proliferation and cancer. *Cell* **162**, 45–58 (2015).
8. Wilson, J. E. et al. Inflammasome-independent role of AIM2 in suppressing colon tumorigenesis via DNA-PK and Akt. *Nat. Med.* **21**, 906–913 (2015).
9. Zaki, M. H. et al. The NOD-like receptor NLRP12 attenuates colon inflammation and tumorigenesis. *Cancer Cell* **20**, 649–660 (2011).

10. Harton, J. A., Linhoff, M. W., Zhang, J. & Ting, J. P. Cutting edge: CATERPILLER: a large family of mammalian genes containing CARD, pyrin, nucleotide-binding, and leucine-rich repeat domains. *J. Immunol.* **169**, 4088–4093 (2002).
11. Conti, B. J. *et al.* CATERPILLER 16.2 (CLR16.2), a novel NBD/LRR family member that negatively regulates T cell function. *J. Biol. Chem.* **280**, 18375–18385 (2005).
12. Liu, R. *et al.* Expression profile of innate immune receptors, NLRs and AIM2, in human colorectal cancer: correlation with cancer stages and inflammasome components. *Oncotarget* **6**, 33456–33469 (2015).
13. Schneider, M. *et al.* The innate immune sensor NLRC3 attenuates Toll-like receptor signaling via modification of the signaling adaptor TRAF6 and transcription factor NF- κ B. *Nat. Immunol.* **13**, 823–831 (2012).
14. Zhang, L. *et al.* NLRC3, a member of the NLR family of proteins, is a negative regulator of innate immune signaling induced by the DNA sensor STING. *Immunity* **40**, 329–341 (2014).
15. Man, S. M. & Kanneganti, T. D. Converging roles of caspases in inflammasome activation, cell death and innate immunity. *Nat. Rev. Immunol.* **16**, 7–21 (2016).
16. Shimobayashi, M. & Hall, M. N. Making new contacts: the mTOR network in metabolism and signalling crosstalk. *Nat. Rev. Mol. Cell Biol.* **15**, 155–162 (2014).
17. Alessi, D. R. *et al.* Characterization of a 3-phosphoinositide-dependent protein kinase which phosphorylates and activates protein kinase B α . *Curr. Biol.* **7**, 261–269 (1997).
18. Stephens, L. *et al.* Protein kinase B kinases that mediate phosphatidylinositol 3,4,5-trisphosphate-dependent activation of protein kinase B. *Science* **279**, 710–714 (1998).
19. Stokoe, D. *et al.* Dual role of phosphatidylinositol-3,4,5-trisphosphate in the activation of protein kinase B. *Science* **277**, 567–570 (1997).
20. Sancak, Y. *et al.* Ragulator–Rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids. *Cell* **141**, 290–303 (2010).
21. Zoncu, R. *et al.* mTORC1 senses lysosomal amino acids through an inside-out mechanism that requires the vacuolar H⁺-ATPase. *Science* **334**, 678–683 (2011).
22. Monick, M. M. *et al.* Lipopolysaccharide activates Akt in human alveolar macrophages resulting in nuclear accumulation and transcriptional activity of β -catenin. *J. Immunol.* **166**, 4713–4720 (2001).
23. Jones, B. W., Heldwein, K. A., Means, T. K., Saukkonen, J. J. & Fenton, M. J. Differential roles of Toll-like receptors in the elicitation of proinflammatory responses by macrophages. *Ann. Rheum. Dis.* **60** (Suppl. 3), iii6–iii12 (2001).
24. Ojaniemi, M. *et al.* Phosphatidylinositol 3-kinase is involved in Toll-like receptor 4-mediated cytokine expression in mouse macrophages. *Eur. J. Immunol.* **33**, 597–605 (2003).
25. Laudien, M. *et al.* Molecular signatures of a disturbed nasal barrier function in the primary tissue of Wegener's granulomatosis. *Mucosal Immunol.* **4**, 564–573 (2011).
26. Shiau, C. E., Monk, K. R., Joo, W. & Talbot, W. S. An anti-inflammatory NOD-like receptor is required for microglia development. *Cell Reports* **5**, 1342–1352 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Transgenic Gene Knockout Shared Resource at St. Jude Children's Research Hospital (SJCRH) for assistance with knockout mouse generation. Images were acquired at the SJCRH Cell & Tissue Imaging Center, which is supported by SJCRH and NCI grant P30 CA021765-35. Work from our laboratory is supported by the US National Institutes of Health (grants AI101935, AI124346, AR056296 and CA163507 to T.-D.K.), ALSAC (to T.-D.K.), and the ExC306 Inflammation at Interfaces, the DFG SFB 877 B9 and DFG SFB1182 C2 projects (to P.R.). S.M.M. is supported by the R. G. Menzies Early Career Fellowship from the National Health and Medical Research Council of Australia.

Author Contributions R.K., S.M.M. and T.-D.K. conceptualized the study; R.K., S.M.M., R.K.S.M. and S.K. designed the methodology; R.K., S.M.M., R.K.S.M., S.K., Q.Z., B.R.S., A.R.B., X.Q., S.P. and P.V. performed the experiments; R.K., S.M.M., R.K.S.M., S.K., Q.Z. and P.V. conducted the analysis; R.K., S.M.M. and T.-D.K. wrote the manuscript; P.R. and T.-D.K. provided resources; T.-D.K. provided overall supervision.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.-D.K. (thirumala-devi.kanneganti@stjude.org).

METHODS

Mice. Wild-type (C57BL/6) and *Nlr3*^{-/-} mice were bred and maintained under specific pathogen-free conditions at St. Jude Children's Research Hospital. To generate the *Nlr3*^{-/-} mice, exons 2 and 3 of the gene encoding NLRC3 (2.5 kb) were excised, which resulted in deletion of 661 amino acids encoding the N-terminal caspase-recruitment domain (also known as CARD) and the central nucleotide-binding domain (also known as NACHT) of NLRC3. Splicing of exon 1 and exon 4 led to a frameshift, generating a premature stop codon (Extended Data Fig. 1c). The targeting vector ROS1-HR was linearized and transfected into embryonic stem (ES) cells. Cells positive for the targeting vector were selected by resistance to 200 µg ml⁻¹ of G418 and further screened for 5' and 3' homologous recombination. ES cell clones with the correct targeting events were used for blastocyst injection. CMV-Cre mice were used to delete the conditional floxed *Nlr3* allele to generate the *Nlr3*^{-/-} mice. B6.129P2-*Lyz2*^{tm1(Cre)Jfo}/J (004781, The Jackson Laboratory), B6.Cg-Tg(Vav1-cre)A2Kio/J (008610, The Jackson Laboratory) and B6.Cg-Tg(Vil-cre)997Gum/J mice (004586, The Jackson Laboratory) were used to delete the conditional floxed *Nlr3* allele in a cell-type-specific manner. Expression of NLRC3 in wild-type and *Nlr3*^{-/-} mice were confirmed by genomic PCR amplification (Extended Data Fig. 1c). *Nlr3*^{-/-} mice were backcrossed to C57BL/6 for nine generations.

Nlr3^{Δ1d} (*Nlr3* large deletion) mice were generated by injection of two single guide RNAs (*Nlr3*-Guide-01: 3'-ATTCCCAGTCTCGTCTTAGGC-5'; 125 ng µl⁻¹) and *Nlr3*-Guide-02: 3'-AGTGGAAACAGCACAGTTCGC-5'; 125 ng µl⁻¹), designed to introduce DNA double-strand breaks into intron 1 and intron 3 of the *Nlr3* gene, together with a codon-optimized Cas9 mRNA transcript (50 ng µl⁻¹) into the cytoplasm of the pronuclear stage C57BL/6J zygotes (Transgenic Core Unit, St. Jude Children's Research Hospital). Injected embryos were surgically transplanted into oviducts of pseudo pregnant CD1 females and new-born mice bearing the intron-1-to-intron-3 deletion (~3.8 kb) were identified by the amplification of a 1.1-kb fragment using primers flanking the two break sites: *Nlr3*-F51: 3'-AGAGTGGTGCCATCTTCTGC-5' and *Nlr3*-R32: 3'-CTCAAGTCAGGGCAGCATGA-5'. Sanger sequencing of the ~1.1-kb amplicon confirmed proper deletion of the 3.8-kb fragment containing exon 2 and 3. The sgRNAs and Cas9 mRNA transcript were designed and generated as described previously²⁷. Potential off-target sites were identified using Cas-OFFinder²⁸, amplified by PCR and sequenced. No off-target site cleavage was observed. Two founder mice were used to establish the mouse lines. Animal study protocols were approved by the St. Jude Children's Research Hospital Committee on Use and Care of Animals.

Azoxymethane-DSS model of colorectal tumorigenesis. Male and female mice were used at the age of six weeks. For cohousing experiments, wild-type and *Nlr3*^{-/-} mice were co-housed for three weeks and separated before injection of azoxymethane. We also performed experiments whereby wild-type and *Nlr3*^{-/-} mice were co-housed for three weeks before injection of azoxymethane and remained co-housed over the course of the experiments. In both cases, the results did not differ. Mice were injected intraperitoneally with 10 mg of azoxymethane (A5486, Sigma) per kg body weight, according to previously established protocols^{7,29}. After 5 days, 2% DSS (9011-18-1, Affymetrix eBioscience) was given in the drinking water for 6 days followed by regular drinking water for 2 weeks. This cycle was repeated twice more with 1.5% DSS and mice were killed on day 80 (Extended Data Fig. 1a). For day 8 samples, mice were injected with azoxymethane, and after 5 days, fed with 2% DSS for 3 days before being killed. For Day 14 samples, mice were injected with azoxymethane, and after 5 days, fed with 2% DSS for 6 days. Mice were then fed with regular water for 3 days and collected. Bone-marrow chimaera studies were performed as described previously⁷. No randomization or blinding was performed.

***Apc*^{Min} model of colorectal tumorigenesis.** C57BL/6J-*Apc*^{Min/+}/J mice (002020, The Jackson Laboratory) were crossed with *Nlr3*^{-/-} mice. Littermate *Apc*^{Min/+} and *Apc*^{Min/+} *Nlr3*^{-/-} mice were administered with either 40 mg kg⁻¹ body weight of the dual inhibitor of PI3K and mTOR, NVP-BEZ235 (N-4288, LC Laboratories) dissolved in 10% (v/v) 1-methyl-2-pyrrolidone (328634, Sigma) plus 90% (v/v) polyethylene glycol 300 (90878, Sigma) or the control vehicle 10% (v/v) 1-methyl-2-pyrrolidone plus 90% (v/v) polyethylene glycol 300 by daily oral gavage for 40 days from 6 weeks of age.

Histology and microscopy analysis. Colons were rolled into a 'Swiss roll' and fixed in 10% formalin, processed and embedded in paraffin by standard techniques. Longitudinal sections of 5 µm were stained with haematoxylin and eosin and examined by a pathologist blinded to the experimental groups. Colitis scores were assigned based on inflammation, ulceration, hyperplasia and the extent or severity of damage. Severity scores for inflammation were assigned as follows: 0 = normal (within normal limits); 2 = minimal (mixed inflammation, small, focal or widely separated, limited to lamina propria); 15 = mild (multifocal mixed inflammation, often extending into submucosa); 40 = moderate (large multifocal lesions within

mixed inflammation involving mucosa and submucosa); 80 = marked (extensive mixed inflammation with oedema and erosions; 100 = severe (diffuse inflammation with transmural lesions and multiple ulcers). Scores for ulceration were assigned as follows: 0 = normal (none); 2 = minimal (only one small focus of ulceration involving fewer than 5 crypts); 15 = mild (a few small ulcers up to 5 crypts); 40 = moderate (multifocal ulcers up to 10 crypts); 80 = marked (multifocal to coalescing ulcers involving more than 10 crypts each); 100 = severe (extensive to diffuse with multiple ulcers covering more than 20 crypts each). Scores of hyperplasia were assigned as follows: 0 = normal; 2 = minimal (some areas with crypts elongated and increased mitoses); 15 = mild (multifocal areas with crypts elongated up to twice the normal thickness, normal goblet cells present); 40 = moderate (extensive areas with crypts up to twice the normal thickness, reduced goblet cells); 80 = marked (mucosa over twice the normal thickness, hyperchromatic epithelium, reduced or rare goblet cells, possibly foci of arborization); 100 = severe (mucosa twice the normal thickness, marked hyperchromasia, crowding or stacking, absence of goblet cells, high mitotic index and arborization). Scores of extent were assigned as follows: 0 = normal (rare or inconspicuous lesions); 2 = minimal (less than 5% involvement); 15 = mild (multifocal but conspicuous lesions, 5–10% involvement); 40 = moderate (multifocal, prominent lesions, 10–50% involvement); 80 = marked (coalescing to extensive lesions or areas of inflammation with some loss of structure, 50–90% involvement); 100 = severe (diffuse lesion with effacement of normal structure, over 90% involvement). The proliferating cells in the intestinal epithelium were detected by immunoperoxidase staining for Ki67 and PCNA. The immunohistochemistry antibodies used were: Ki67 (NBPI-40684, Novus), PCNA (M0879, DAKO), β-catenin (610154, BD), P-AKT Ser473 (4060, Cell Signaling), and P-S6 Ribosomal Protein Ser235-236 (4858, Cell Signaling). Tissues were counterstained with haematoxylin. The number of Ki67⁺ or PCNA⁺ cells per crypt in each animal was counted (at least 18–20 crypts per mouse).

Cell culture and stimulation of cells. Pinna of adult wild-type and *Nlr3*^{-/-} mice were minced and digested with 100 mg ml⁻¹ collagenase type IV (LS004188, Worthington Biochemical Corporation) for 3 h, followed by filtration through 70-µm strainers to obtain fibroblasts. Cells were cultured in 50% FBS (TMS-013-B, Millipore) in DMEM (11995073, ThermoFisher Scientific) supplemented with HEPES (15630-080, ThermoFisher Scientific), 1% penicillin and streptomycin (15070-063, ThermoFisher Scientific), L-glutamine (25030, ThermoFisher Scientific), sodium pyruvate (11360, ThermoFisher Scientific), non-essential amino acids (11140, ThermoFisher Scientific), and β-mercaptoethanol (21985023, ThermoFisher Scientific) for the first 3–4 days. Cells were subcultured in DMEM supplemented with 10% FBS and 1% penicillin and streptomycin. All primary fibroblasts were used before reaching the sixth passage. Fibroblasts were seeded onto six-well plates at a density of 2 × 10⁵ cells per well. Cells were deprived of serum for 36 h and further incubated in PBS for 1 h. Cells were then stimulated with 50 ng ml⁻¹ of recombinant murine IGF-1 (250-19, Peprtech) for the indicated time.

BMDMs were cultured as described previously³⁰. BMDMs were stimulated with 500 ng ml⁻¹ ultrapure LPS from *Salmonella minnesota* R595 (tlrl-smpls, InvivoGen) for the indicated time. The human colorectal carcinoma HCT116 cell line (ATCC#CCL-247, American Type Culture Collection) was cultured in McCoy's 5A medium (16600-082, ThermoFisher Scientific) supplemented with 10% FBS and 1% penicillin and streptomycin. The embryonic kidney epithelial cell line HEK293T (ATCC#3216, American Type Culture Collection) and L929 cell line (ATCC#CRL-2648, American Type Culture Collection) were cultured in DMEM supplemented with 10% FBS. All cell lines were maintained at 37 °C with 5% CO₂.

Colon organoid culture. Mouse colon stem cells were cultured using IntestiCult organoid growth medium according to the manufacturer's instructions (06005, STEMCELL Technologies). The whole colon was removed from untreated wild-type and *Nlr3*^{-/-} mice and rinsed with cold PBS. The colon was cut into 2-mm segments and washed 20 times with cold PBS. Colonic segments were incubated in Gentle Cell Dissociation Reagent (07174, STEMCELL Technologies), rotated at 350g for 15 min at room temperature, followed by re-suspension in PBS supplemented with 0.1% BSA (A6003, Sigma). Dissociated colonic crypts were filtered through 70-µm strainers. Dissociated colonic crypts were resuspended in DMEM/F12 medium with 15 mM HEPES (36254, STEMCELL Technologies), counted and resuspended in IntestiCult organoid growth medium and Matrigel (356230, Corning) in a 1:1 ratio. Cells were plated in 24-well culture plates (3738, Corning). IntestiCult organoid growth medium were added to the cell culture plates to immerse the matrix composed of IntestiCult organoid growth medium and Matrigel. For inhibition studies, 50 µM of LY294002 (440202, Millipore) or 10 µg ml⁻¹ rapamycin (553210, Sigma) was added to the IntestiCult organoid growth medium.

Proliferation assay. Cell proliferation was measured using the WST-1 reagent (05015944001, Roche). Primary ear fibroblasts or HCT116 cells were plated at a density of 5,000 cells per well in 96-well tissue culture plates and incubated

overnight. Cells were deprived of serum for 36 h and exposed to normal culture media for the indicated time interval. The WST-1 reagent was added to the cells for 2 h. Plates were then read at 450 nm. The number of cells was calculated using a standard curve³¹.

Immunoblotting. Proteins were extracted from colon tissue or cells using RIPA lysis buffer supplemented with protease (11697498001, Roche) and phosphatase inhibitors (04906837001, Roche) as described previously³². Samples were resolved in 8–15% SDS–PAGE and transferred onto PVDF membranes (IPVH00010, Millipore). Blocking was performed in 5% milk for 1 h and membranes were incubated in primary antibodies overnight at 4 °C. Membranes were incubated with HRP-conjugated secondary antibody for 1 h and proteins were visualized using Super Signal Femto substrate (34096, ThermoFisher Scientific). The primary antibodies used were P-ERK (1:1,000, 9101, Cell Signaling), ERK (1:1,000, 9102, Cell Signaling), P-IκBα (1:1,000, 9241, Cell Signaling), IκBα (1:1,000, 9242, Cell Signaling), P-AKT Ser473 (1:1,000, 4060, Cell Signaling), P-AKT Thr308 (1:1,000, 13038), AKT (1:1,000, 4691, Cell Signaling), P-mTOR Ser2448 (1:1,000, 2971, Cell Signaling), mTOR (1:1,000, 2972, Cell Signaling), P-p70 S6K Thr389 (1:1,000, 9205, Cell Signaling), P-S6 Ser235/236 (1:1,000, 4856, Cell Signaling), P-4E-BP1 Thr37/46 (1:1,000, 2855, Cell Signaling), P-STAT3 Tyr705 (1:1,000, 9131, Cell Signaling), P-PDK1 (1:1,000, 3061, Cell Signaling), PDK1 (1:1,000, 3062, Cell Signaling), P-PI3K p85 (1:1,000, 4228, Cell Signaling), PI3K p85 (1:1,000, 4257, Cell Signaling), PI3K p110α (1:1,000, 4249, Cell Signaling), mouse anti-GFP (1:2,000, sc-9996, Santa Cruz Biotechnology), rabbit anti-GFP (1:2,000, sc-8334, Santa Cruz Biotechnology), β-actin (1:2,000, 8457, Cell Signaling), and anti-GAPDH (1:10,000, 5174, Cell Signaling). Immunoblots were quantified using ImageJ.

Immunofluorescence staining. Serum-deprived primary ear fibroblasts were left untreated or stimulated with IGF-1 for 30 min. Fibroblasts or six-day-old intestinal organoids were washed three times with PBS and were fixed for 15 min at room temperature in 4% paraformaldehyde, followed by blocking in 10% normal goat serum (X090710-8, Dako) supplemented with 0.1% saponin (47036, Sigma). Fibroblasts or organoids were incubated with the following antibodies overnight at 4 °C: P-AKT Ser473 (1:200, 4060, Cell Signaling), P-AKT Thr308 (1:200, 13038); P-S6 Ser235/236 (1:500, 4856, Cell Signaling), P-4E-BP1 Thr37/46 (1:200, 2855, Cell Signaling). Samples were also stained with Alexa Fluor 488 phalloidin (1:500, A12379, ThermoFisher Scientific) or Alexa Fluor 488 E-cadherin (1:200, 53-3249-80, Affymetrix eBioscience). To analyse mTOR activation, cells were incubated overnight at 4 °C with antibodies against mTOR (1:200, 2983, Cell Signaling) and LAMP1 (1:1,000, eBioD4B, Affymetrix eBioscience). The secondary antibodies used were Alexa Fluor 568-conjugated antibody to rabbit immunoglobulin G (1:250; A11036; ThermoFisher Scientific), Alexa Fluor 568-conjugated antibody to rat immunoglobulin G (1:250; A11077; ThermoFisher Scientific) and Alexa Fluor 488-conjugated antibody to rabbit immunoglobulin G (1:250; A11034; ThermoFisher Scientific). Cells and organoids were counterstained in DAPI mounting medium (H-1200, Vecta Labs) and images taken with a Nikon C2 confocal microscope. The average density unit and the Pearson's correlation coefficient were calculated using the digital microscopy imaging software SlideBook 5 (Intelligent Imaging Innovations).

Quantitative reverse-transcriptase PCR. RNA was isolated using TRIzol (15596026, ThermoFisher Scientific) and converted to cDNA using the High-Capacity cDNA Reverse Transcription kit (4368814, Applied Biosystems). Gene expression was assessed using the 2 × SYBR Green kit (4368706, Applied Biosystems) according to the manufacturer's instructions. Primer sequences are listed in the Supplementary Information.

Cytokine measurement by ELISA. Cytokines in the colon and sera were measured by ELISA according to manufacturer's instructions. IL-18 was measured using an ELISA kit (BMS618/3TEN, Affymetrix eBioscience) and all other cytokines were measured by a multiplex ELISAs (MCYTOMAG-70K, Millipore).

Flow cytometry. Colons were dissected, washed with ice-cold PBS and cut into small pieces. Colons pieces were incubated with PBS containing 1 mM DTT, 5 mM EDTA and 10 mM HEPES at 37 °C for 30 min with gentle shaking to remove the epithelial layer. The colon segments were further digested in RPMI medium containing 0.5 mg ml⁻¹ collagenase D at 37 °C for 1.5 h. The supernatant was passed through 70-μm cell strainer and enriched by 37.5% Percoll to isolate lamina propria cells.

The following monoclonal antibodies were used for flow cytometry: CD4 (RM4-5; 14-0042-85), CD11b (M1/70; 48-0112-82) and CD8a (53-6.7; 48-0081-82) from Affymetrix eBioscience, CD19 (6D5; 115512), NK1.1 (PK136; 108708), CD11c (N418; 117306), Gr1 (RB6-8C5; 108426) and F4/80 (BM8; 123109) from BioLegend. The dilution factor for all antibodies was 1:300. The following gating strategies were used: B cells were gated as live cells and CD19⁺. CD4⁺ T cells were gated as live cells, CD4⁺ and CD8⁻. CD8⁺ T cells were gated as live cells, CD8⁺ and CD4⁻. Natural killer cells were gated as live cells and NK1.1⁺. Macrophages were gated as live cells, CD11b⁺, Gr1^{low-neg}, F4/80⁺ and CD11c⁻. Neutrophils were gated as live cells, CD11b⁺ and Gr1^{hi}. CD11b⁺CD11c⁺ cells were gated as live cells,

CD11b⁺, Gr1^{low-neg}, CD11c⁺ and F4/80⁻. Flow cytometry data were acquired on a BD FACSCalibur and analysed using TreeStar FlowJo software.

T-cell stimulation and intracellular cytokine staining. Splenocytes from wild-type and *Nlr3*^{-/-} mice were treated with ACK lysis buffer at room temperature for 1 min to remove red blood cells. Splenocytes were washed, counted and plated at 2 × 10⁵ cells per well in a 96-well plate coated with 1 μg ml⁻¹ anti-CD3 (145-2C11, Affymetrix eBioscience) and 1 μg ml⁻¹ anti-CD28 (16-0281, Affymetrix eBioscience). Cells were cultured at 37 °C in the presence and absence of 20 ng ml⁻¹ murine IL-2 (212-12, Peprotech) for 4 days. Brefeldin A (00-4506, Affymetrix eBioscience) was added to the media for 3 h, followed by washing in PBS, and staining with anti-CD4 (14-0042-85, Affymetrix eBioscience) and anti-CD3 (145-2C11, Affymetrix eBioscience) antibodies on ice for 20 min. Stained cells were fixed in 1% paraformaldehyde for 30 min on ice and permeabilized using permeabilization buffer (00-8333-56, Affymetrix eBioscience) according to manufacturer's instructions. To detect intracellular cytokines, fixed cells were stained with anti-IFNγ (50-7311, Tonbo) and anti-TNF (506322, Biolegend) for 30 min on ice. Flow cytometry were performed as described above.

siRNA knockdown. Primary ear fibroblasts were transfected with a siRNA from siGENOME smart pools with the assistance of the Neon Transfection System (MPK5000, ThermoFisher Scientific). The siGENOME SMARTpool siRNA specific for the gene encoding mouse NLRC3 (M-052823-01, Dharmacon) and a control siRNA pool were used. Sequences for siRNA are listed in Supplementary Information. After 48 h of transfection, cells were stimulated with IGF-1 as described above.

Retroviral transduction. Human or mouse MSCV-NLRC3-IRES-GFP or MSCV-IRES-GFP construct was co-transfected with retroviral packaging plasmids (pPAM-E and VSV-G) into HEK293T cells using Xfect Transfection Reagents (631318, Clontech Laboratories, Inc.). Virus-containing media were collected 48–72 h later and passed through a 0.45-μm filter. Primary ear fibroblasts or HCT116 cells were transduced with control or NLRC3-encoding retroviral vectors. Cells expressing GFP were selected by flow cytometry.

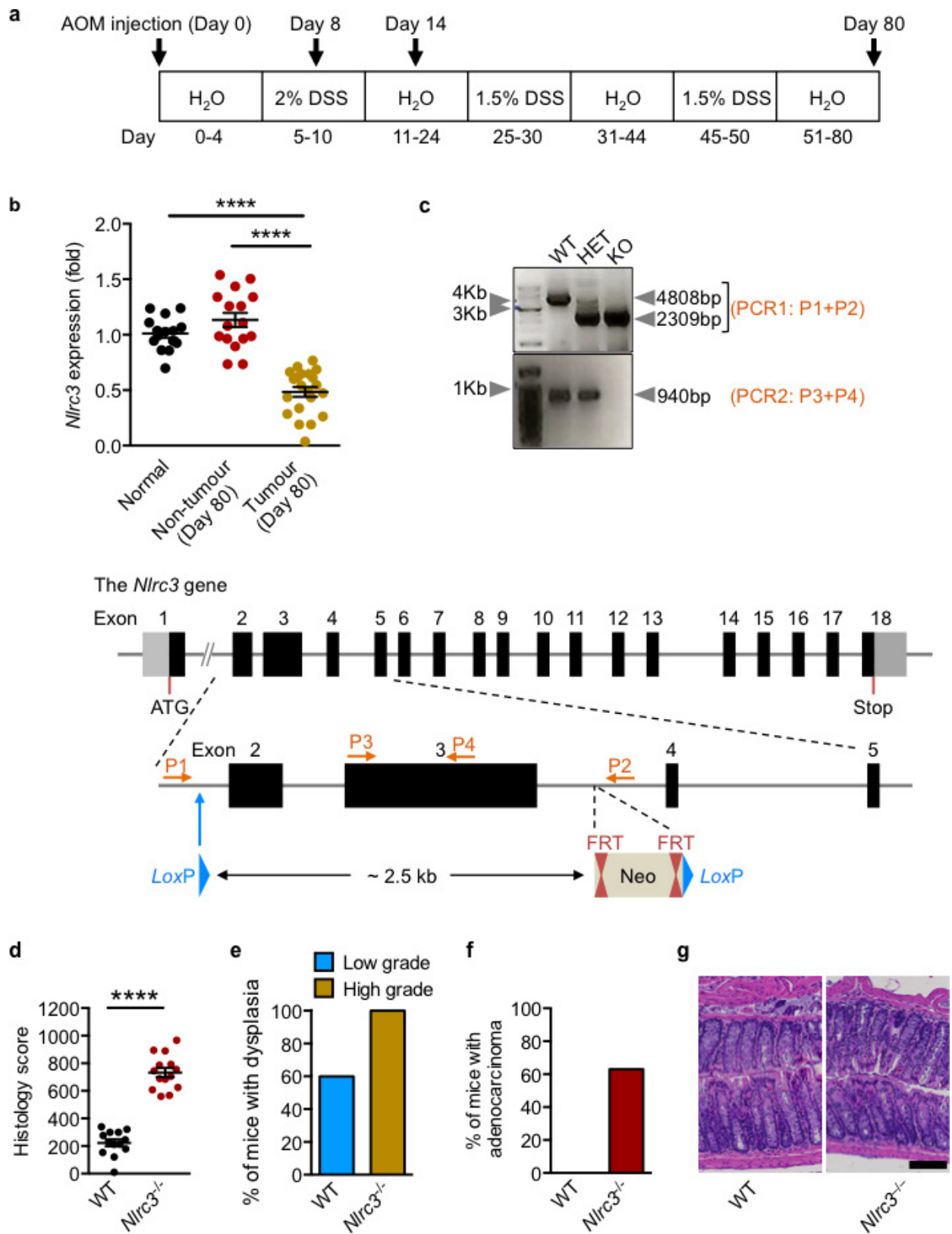
Generation of cells expressing NLRC3 and its deletion mutants. Plasmids pVSVg, pPAM-E and pMIGII encoding the mouse *Nlr3* gene or the mouse *Nlr3* gene lacking regions encoding the CARD, NACHT or LRR domain was transfected into L929 cells to generate retroviral stocks. Domains were annotated in accordance with the NCBI Conserved Domain (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Retroviral supernatants were collected after 48 h of transfection and filtered through 0.4-μm filters. L929 cells were infected with the retroviral stocks in the presence of polybrene to generate cells that stably expressed either wild-type NLRC3 or NLRC3 lacking regions encoding the CARD, NACHT or LRR domain.

Co-immunoprecipitation. Cells were collected with ice-cold PBS and lysed in lysis buffer composed of 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP-40, protease and phosphatase inhibitors. Lysates were cleared of insoluble material by centrifugation at 15,000g for 10 min. For immunoprecipitation, cell lysates were incubated with 3 μg of primary antibodies at 4 °C for 12–16 h on a rocking platform, followed by incubation with Protein A/G PLUS-Agarose (sc-2003, Santa Cruz) for a further 2 h on a rocking platform. The immunoprecipitated products were washed three times with lysis buffer and eluted using 2 × SDS sample buffer and boiled at 100 °C for 5 min.

Statistical analysis. GraphPad Prism 6.0 software was used for data analysis. Data are shown as mean ± s.e.m. Statistical significance was determined by *t*-test (two-tailed) for two groups or one-way ANOVA (with multiple comparisons tests) for three or more groups. *P* < 0.05 was considered statistically significant. No statistical methods were used to predetermine sample size.

Data availability. Source data for Figs 3a, d, e, h, 4d–f and Extended Data Figs 3b, 5a, 6a, b, d, e, 9a, b, d–g, 10a, b have been provided in Supplementary Fig. 1. All other data supporting the findings of this study are available from the corresponding author on request.

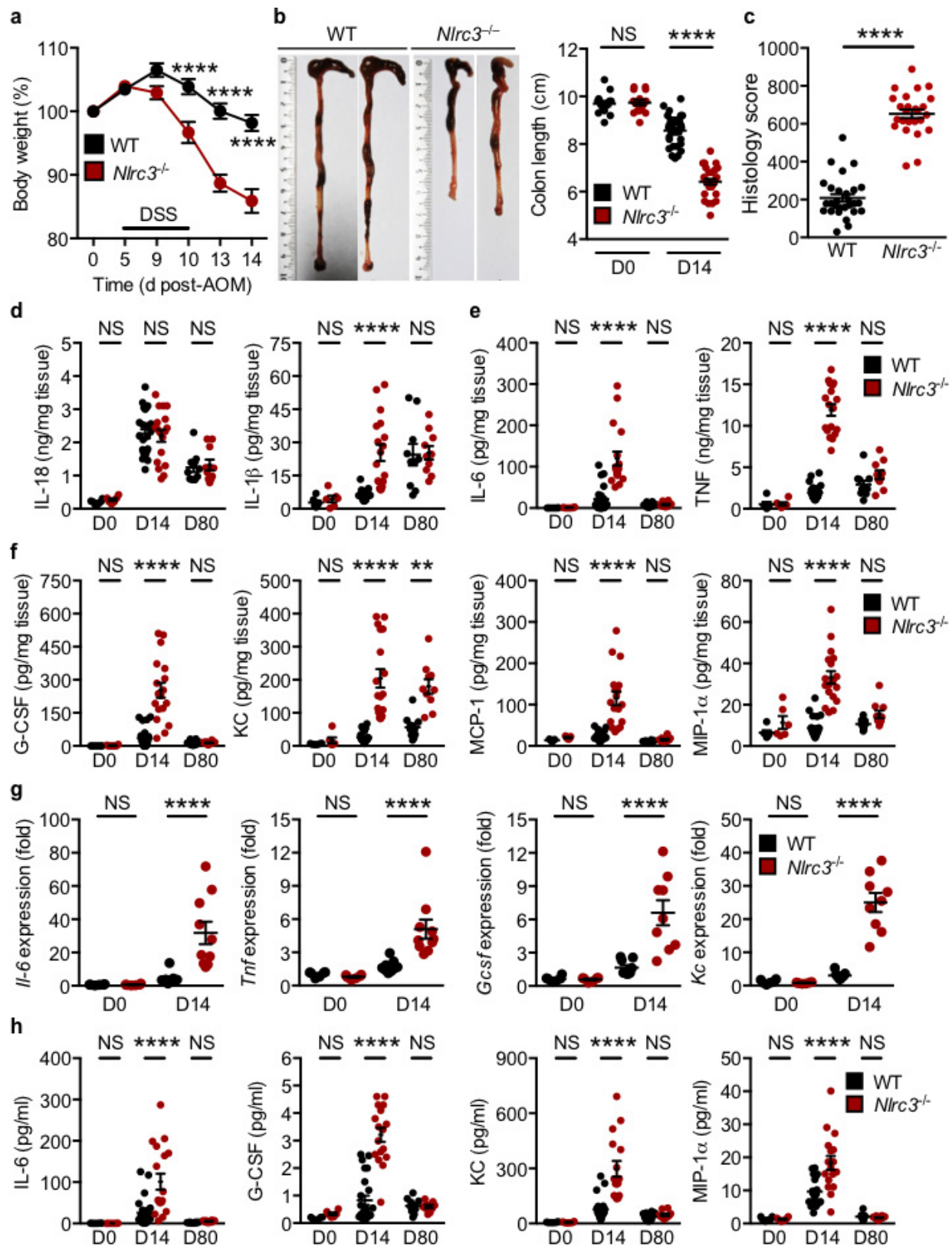
27. Pelletier, S., Gingras, S. & Green, D. R. Mouse genome engineering via CRISPR–Cas9 for study of immune function. *Immunity* **42**, 18–27 (2015).
28. Bae, S., Park, J. & Kim, J. S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
29. Zhu, Q. *et al.* Cutting edge: STING mediates protection against colorectal tumorigenesis by governing the magnitude of intestinal inflammation. *J. Immunol.* **193**, 4779–4782 (2014).
30. Man, S. M. *et al.* The transcription factor IRF1 and guanylate-binding proteins target activation of the AIM2 inflammasome by *Francisella* infection. *Nat. Immunol.* **16**, 467–475 (2015).
31. Boussemer, L. *et al.* eIF4F is a nexus of resistance to anti-BRAF and anti-MEK cancer therapies. *Nature* **513**, 105–109 (2014).
32. Karki, R. *et al.* Concerted activation of the AIM2 and NLRP3 inflammasomes orchestrates host protection against *Aspergillus* infection. *Cell Host Microbe* **17**, 357–368 (2015).



Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | NLRC3 prevents colitis-associated colorectal tumorigenesis. **a**, Timeline for azoxymethane (AOM) and DSS treatment. **b**, Relative expression levels of the gene encoding NLRC3 in tumour and non-tumour tissue in the colon of wild-type mice 80 days after azoxymethane injection. **c**, Targeting strategy used to generate *Nlrc3*^{-/-} mice and PCR analysis for the gene encoding NLRC3 in wild-type C57BL/6 mice, *Nlrc3*^{+/-} mice and *Nlrc3*^{-/-} mice. The primers P1 (which binds a region between exon 1 and exon 2) and P2 (which binds a region between exon 3 and exon 4) were designed for 'PCR1' such that it generates a 4,804-bp PCR fragment for the wild-type allele and a 2,309-bp fragment for the knockout (KO) allele. However, PCR1 cannot differentiate heterozygote (HT) and knockout mice because the knockout

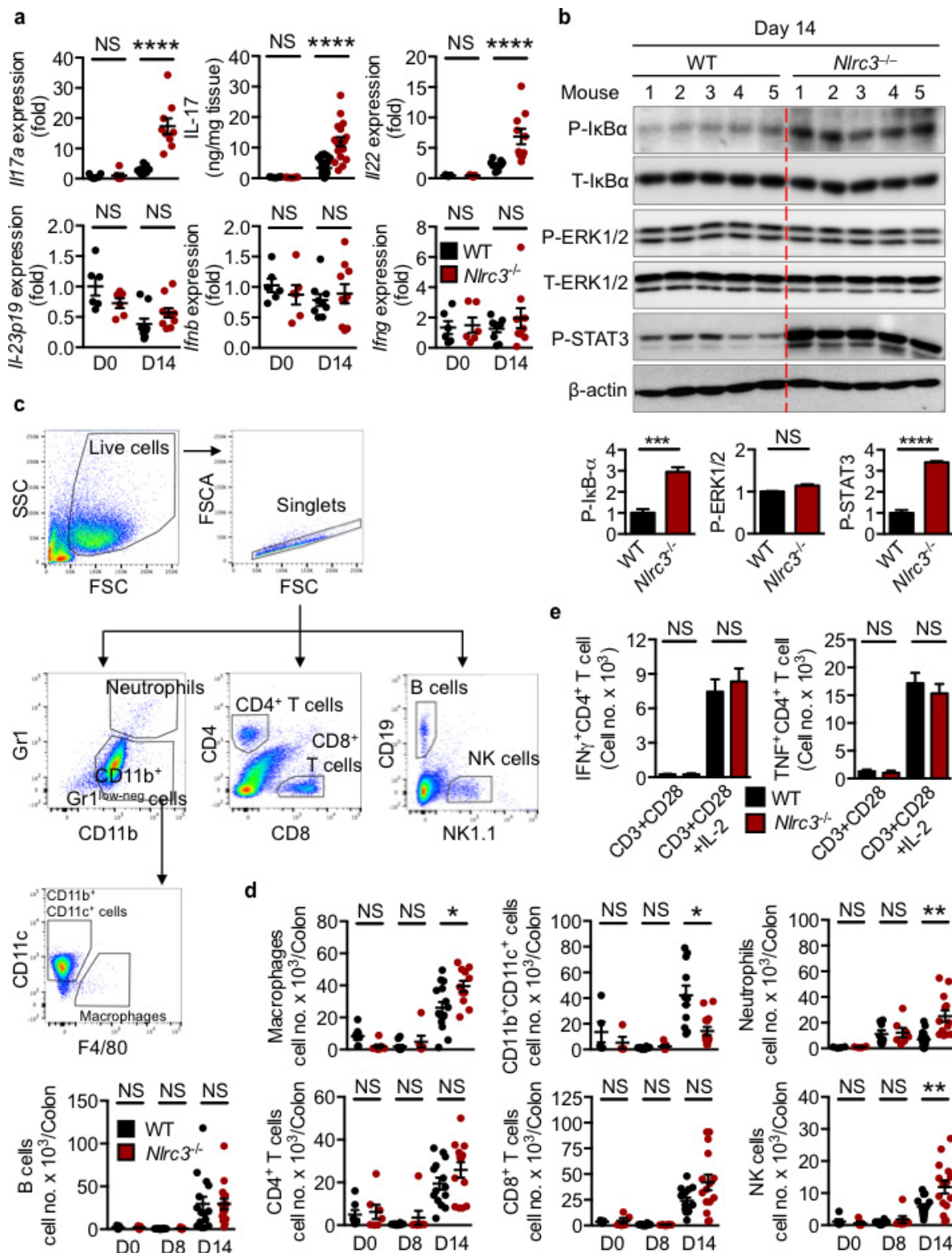
2,309-bp fragment outcompetes the wild-type 4,804-bp fragment. Therefore, we designed primers P3 and P4 for use in PCR2 to amplify a 940-bp fragment from exon 3 to confirm its presence in wild-type and HET mice and its absence in knockout mice. **d**, Histological scores of the colon tissue in wild-type and *Nlrc3*^{-/-} mice 80 days after azoxymethane injection **e**, Percentages of mice with dysplasia 80 days after injection of azoxymethane. **f**, Percentages of mice with adenocarcinoma 80 days after injection of azoxymethane. **g**, Haematoxylin and eosin staining of colon crypts. Scale bar, 100 μ m. Each symbol represents an individual mouse (**b**, **d**). **** $P < 0.0001$; NS, not statistically significant by one-way ANOVA (**b**) or two tailed *t*-test (**d**). Data represent two independent experiments (**b**, **d**-**g**) and are presented as mean \pm s.e.m. (**b**, **d**).



Extended Data Figure 2 | NLRC3 dampens intestinal inflammation.

a, Body-weight change of mice pooled from three independent experiments. **b**, Images of colon and colon length in wild-type mice and *Nlr3*^{-/-} mice at day 14. **c**, Histological scores at day 14. **d**, Levels of IL-18 and IL-1 β in colon tissue at days 14 and 80. **e**, **f**, Levels of IL-6, TNF, G-CSF, KC, MCP-1 and MIP-1 α in colon tissue at days 14 and 80. **g**, Relative expression of genes encoding IL-6, TNF, G-CSF and KC in colon

tissue of untreated wild-type and *Nlr3*^{-/-} mice and in wild-type and *Nlr3*^{-/-} mice at day 14. **h**, Levels of IL-6, G-CSF, KC and MIP-1 α in sera of untreated wild-type and *Nlr3*^{-/-} mice and in wild-type and *Nlr3*^{-/-} mice at days 14 and 80. Each symbol represents an individual mouse (**b–h**). ***P* < 0.01; *****P* < 0.0001; NS, not statistically significant by one-way ANOVA (**a**) or two-tailed *t*-test (**b–h**). Data represent three independent experiments (**a–h**) and are presented as mean \pm s.e.m. in **a–h**.

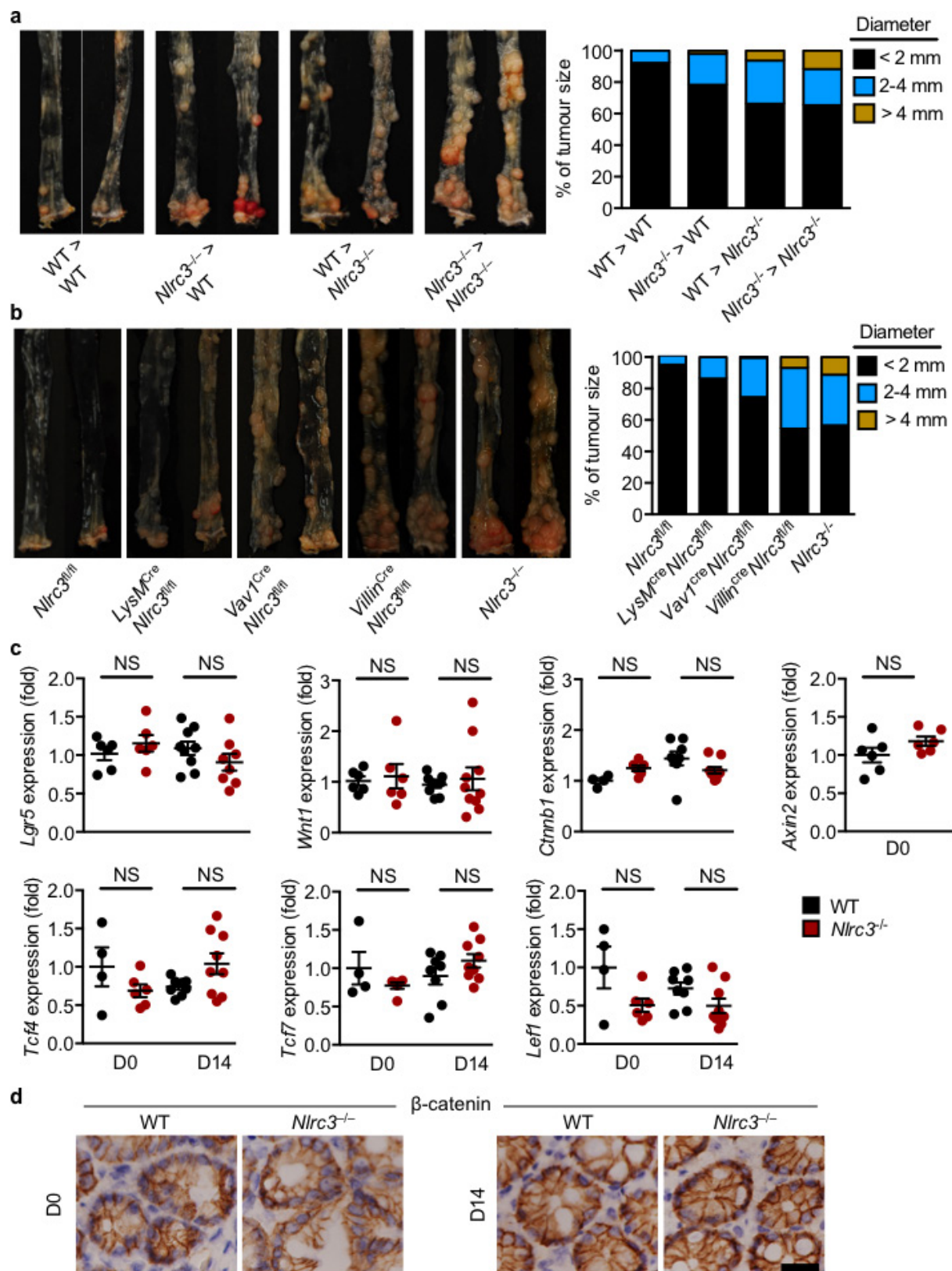


Extended Data Figure 3 | NLRC3 governs colorectal tumorigenic susceptibility via inflammatory mediators and immune cells.

a, Relative expression of genes encoding IL-17a, levels of the IL-17 protein, and relative expression of genes encoding IL-22, IL-23p19, IFN β and IFN γ in colon tissue of untreated wild-type and *Nlr3*^{-/-} mice and in wild-type and *Nlr3*^{-/-} mice at day 14. **b**, Immunoblot analysis of phosphorylated and total I κ B α (P-I κ B α and T-I κ B α), ERK1 and ERK2 (P-ERK1/2 and T-ERK1/2), phosphorylated STAT3 (P-STAT3), and β -actin (loading control) in colon tissue of wild-type and *Nlr3*^{-/-} mice 14 days after injection of azoxymethane (top). The protein band intensity was normalized to the total protein counterpart and/or β -actin, and

expressed relative to that of wild-type controls, set at 1 (bottom).

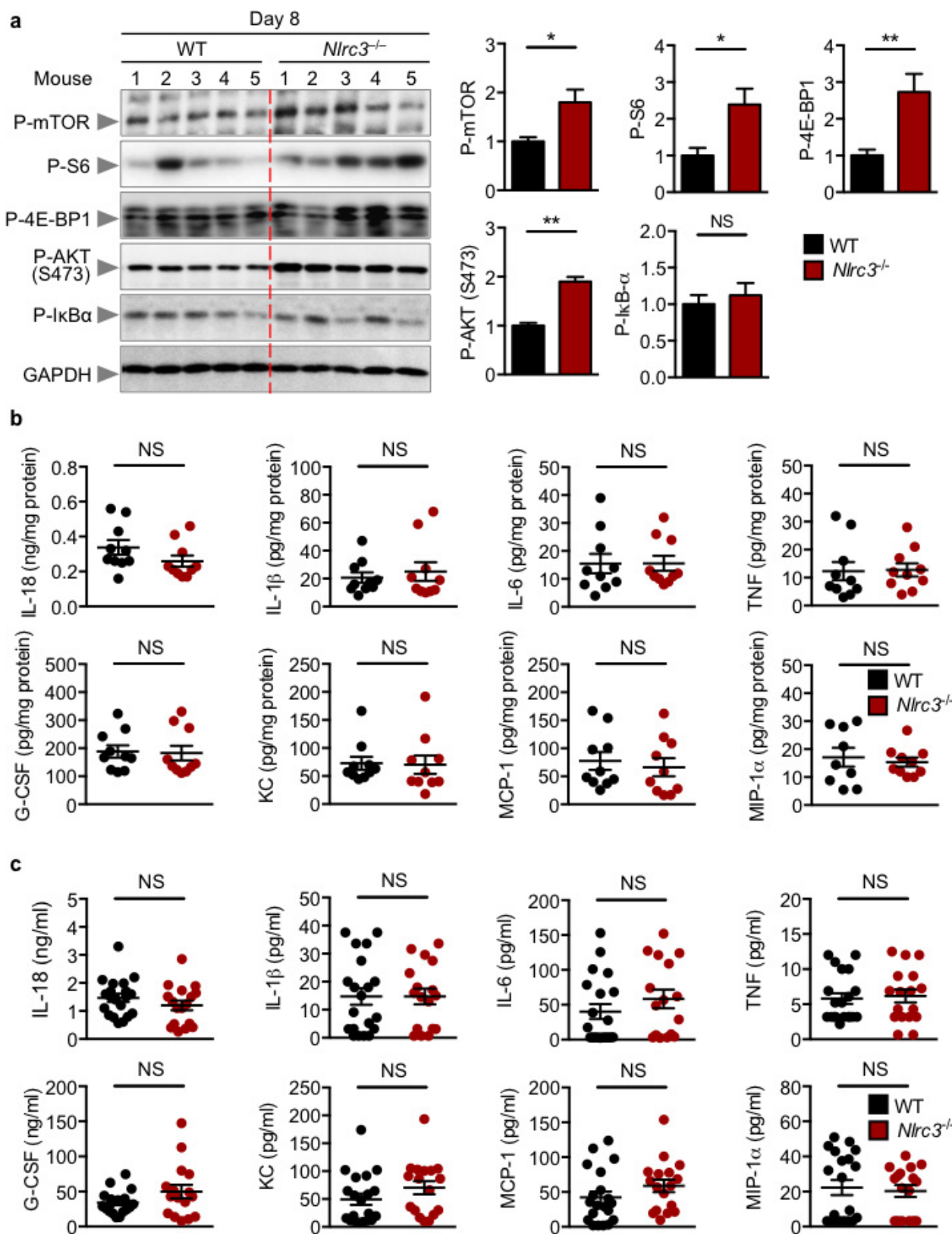
c, Gating strategies used to generate data in **d**. **d**, Number of macrophages, CD11b⁺CD11c⁺ cells, neutrophils, B cells, CD4⁺ T cells, CD8⁺ T cells and natural killer cells per colon in wild-type and *Nlr3*^{-/-} mice at days 8 and 14. **e**, Splenocytes from wild-type and *Nlr3*^{-/-} mice were stimulated with CD3, CD28, and IL-2 and intracellular staining was performed for IFN γ and TNF. * P < 0.05; ** P < 0.01; *** P < 0.001; **** P < 0.0001; NS, not statistically significant by two-tailed t -test (**a**, **b**, **d**, **e**). Data pooled from two independent experiments (**a**) or represent one experiment representative of two independent experiments (**b**–**e**) and are presented as mean \pm s.e.m. in **a**, **b**, **d**, **e**.



Extended Data Figure 4 | The inhibitory effect of NLRC3 is more dominant in intestinal epithelial cells than in haematopoietic cells.

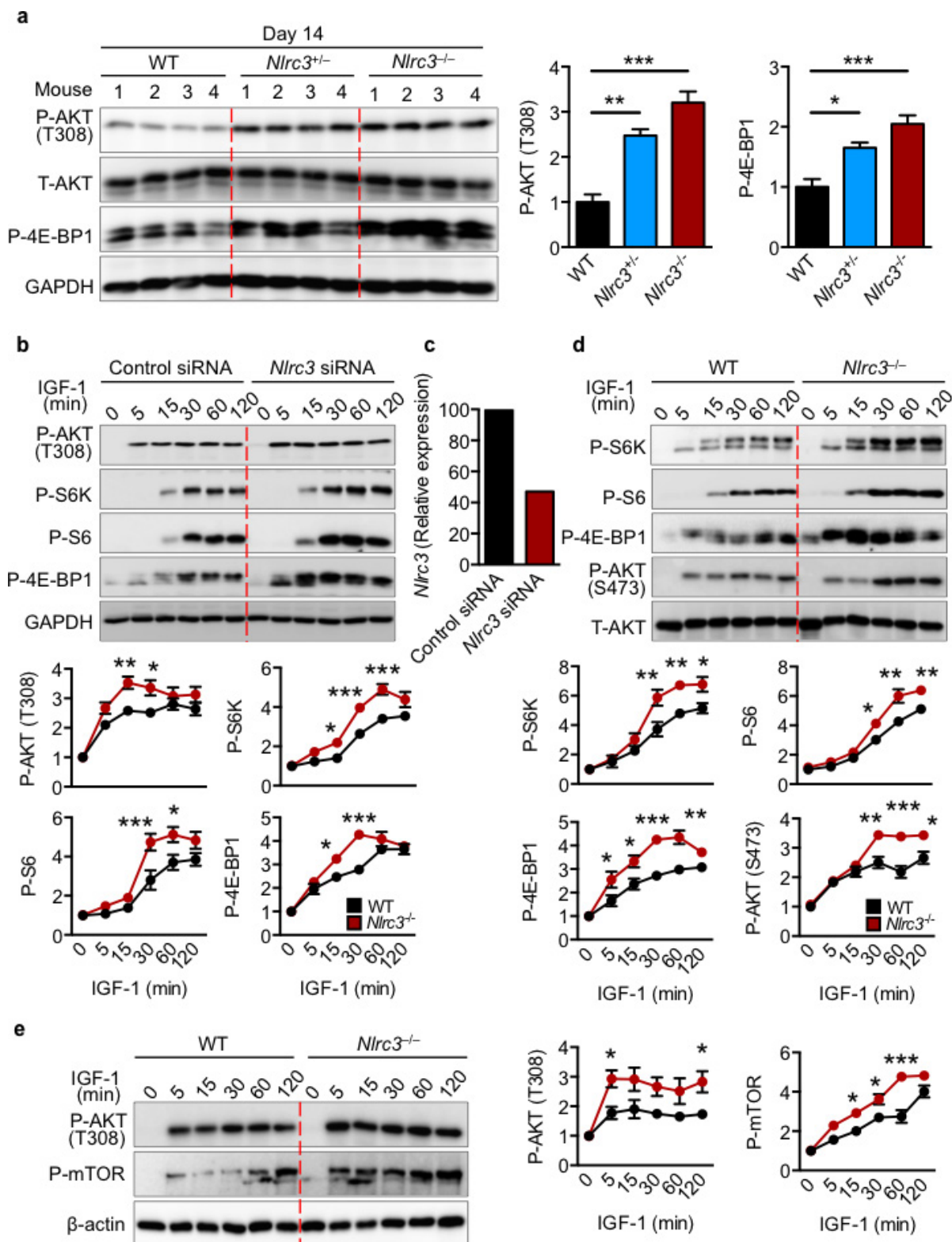
a, Left panel shows colon tumours in, from left to right, wild-type \rightarrow wild-type ($n=10$); *Nlrc3*^{fl/fl} \rightarrow wild-type ($n=9$); wild-type \rightarrow *Nlrc3*^{fl/fl} ($n=8$); *Nlrc3*^{fl/fl} \rightarrow *Nlrc3*^{fl/fl} ($n=9$) bone-marrow chimeric mice at day 80. Percentages of the tumour size of mice are shown on the right. **b**, Left panel shows colon tumours in littermate *Nlrc3*^{fl/fl} ($n=8$), *LysM*^{Cre}*Nlrc3*^{fl/fl} ($n=11$), *Vav1*^{Cre}*Nlrc3*^{fl/fl} ($n=9$), *Villin*^{Cre}*Nlrc3*^{fl/fl} ($n=7$) and *Nlrc3*^{fl/fl} ($n=8$) mice at day 80. The percentages of tumours of each size of each

mouse type are shown on the right. **c**, Relative expression of genes encoding LGR5, WNT1, β -catenin (*Cttnb1*), Axin2, TCF4, TCF7, and LEF1 in colon tissue of untreated wild-type and *Nlrc3*^{fl/fl} mice or in wild-type and *Nlrc3*^{fl/fl} mice at day 14. **d**, Immunohistochemical staining of β -catenin in colon tissue of wild-type and *Nlrc3*^{fl/fl} mice. Scale bar, 20 μ m. Each symbol represents one mouse (**c**). NS, not statistically significant by two-tailed *t*-test. Data represent two independent experiments and are presented as mean \pm s.e.m. in **c**.



Extended Data Figure 5 | Dysregulation of mTOR signalling precedes dysregulation of NF-κB signalling. **a**, Immunoblot analysis of phosphorylated mTOR, S6, 4E-BP1, AKT and IκBα, and GAPDH (loading control) in colon tissue of wild-type and *Nlr3*^{-/-} mice at day 8 (left). The protein band intensity was normalized to GAPDH and expressed as a level relative to that of the wild-type controls, set at 1 (right). **b**, Levels of IL-18,

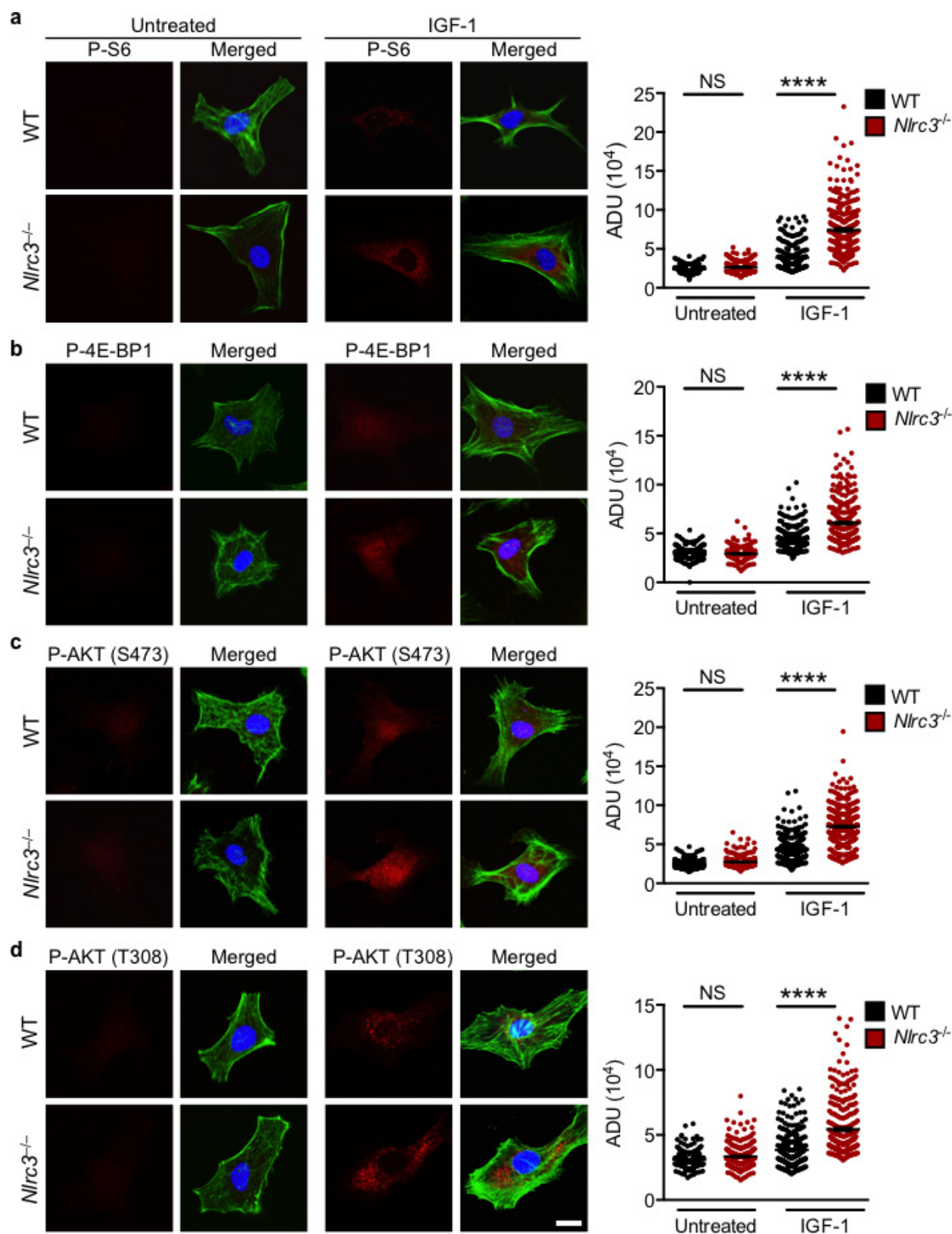
IL-1β, IL-6, TNF, G-CSF, KC, MCP-1 and MIP-1α in colon tissue at day 8. **c**, Levels of IL-18, IL-1β, IL-6, TNF, G-CSF, KC, MCP-1 and MIP-1α in sera. Each symbol represents an individual mouse (**b**, **c**). **P* < 0.05; ***P* < 0.01; NS, not statistically significant by two-tailed *t*-test (**a**–**c**). Data represent two independent experiments and are presented as mean ± s.e.m. in **a**–**c**. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 6 | NLRC3 regulates mTOR activity.

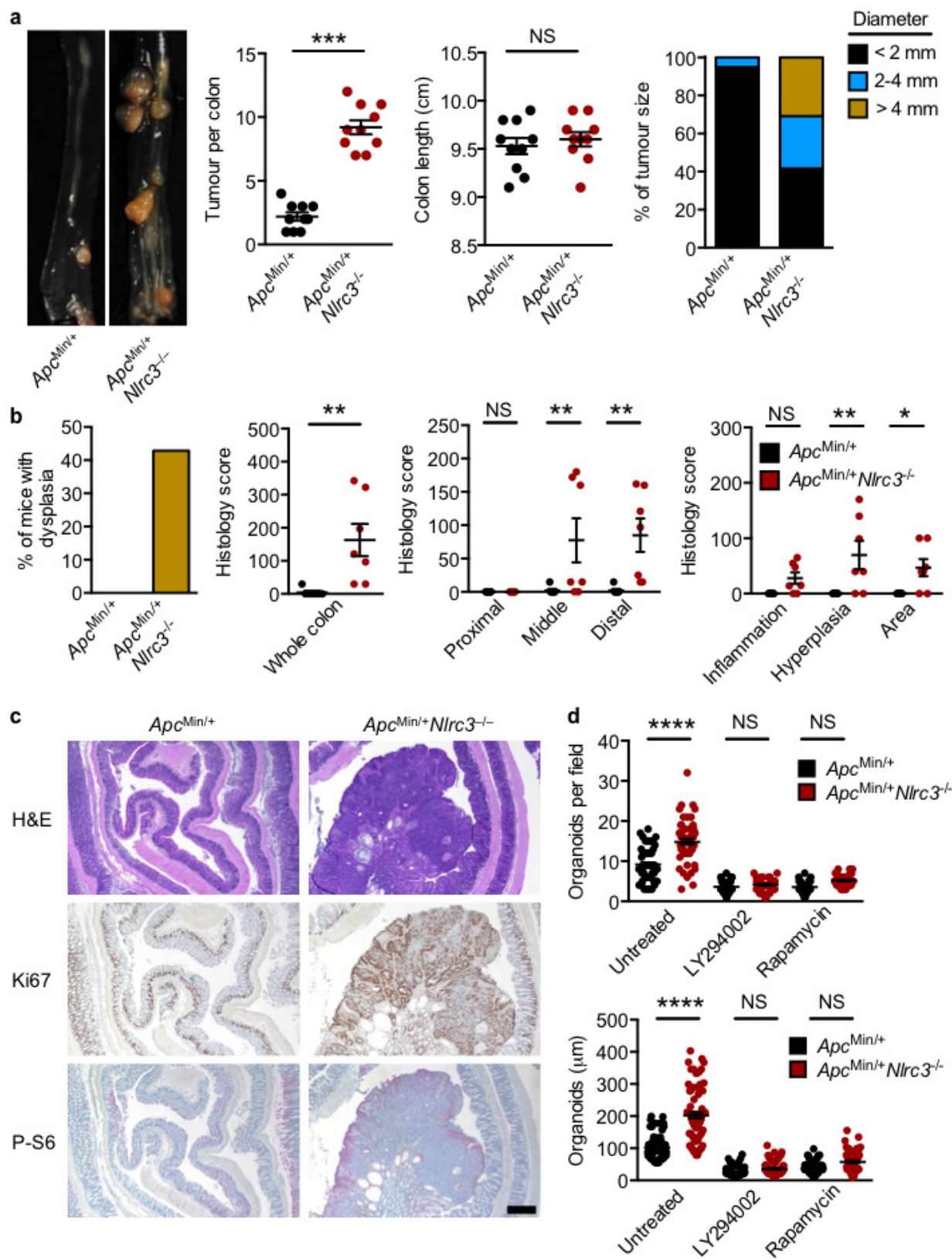
a, Immunoblot analysis of phosphorylated AKT, total AKT, phosphorylated 4E-BP1 and GAPDH (loading control) in the colon tissue of wild-type, *Nlrc3*^{+/-} and *Nlrc3*^{-/-} mice at day 14. The protein-band intensity was normalized to the total protein counterpart and/or the loading control and expressed relative to that of wild-type controls, set at 1 (right). **b**, Immunoblot analysis of phosphorylated AKT, S6K, S6 and 4E-BP1, and GAPDH (loading control) in wild-type fibroblasts transfected with a control siRNA or *Nlrc3* siRNA left untreated or treated with IGF-1 (top). Densitometry analysis as in **a** over 120 min (bottom). **c**, Relative expression of the gene encoding NLRC3 in wild-type fibroblasts

transfected with a control siRNA or compared with wild-type fibroblasts transfected with an *Nlrc3* siRNA. **d**, Immunoblotting of phosphorylated S6K, S6, 4E-BP1, AKT, and total AKT in primary fibroblasts either left untreated or treated with IGF-1 (top). Densitometry analysis as in **a** (bottom). **e**, Immunoblotting of phosphorylated AKT and mTOR, and β-actin (loading control) in primary fibroblasts either left untreated or treated with IGF-1 (left). Densitometry analysis as in **b** (right). **P* < 0.05; ***P* < 0.01; ****P* < 0.001 by one-way ANOVA (**a**, **b**, **d**, **e**). Data are from one experiment representative of two (**a**, **c**) or four independent experiments (**b**, **d** and **e**) and are presented as mean ± s.e.m. in **a**, **b**, **d**, **e**. For gel source data, see Supplementary Fig. 1.



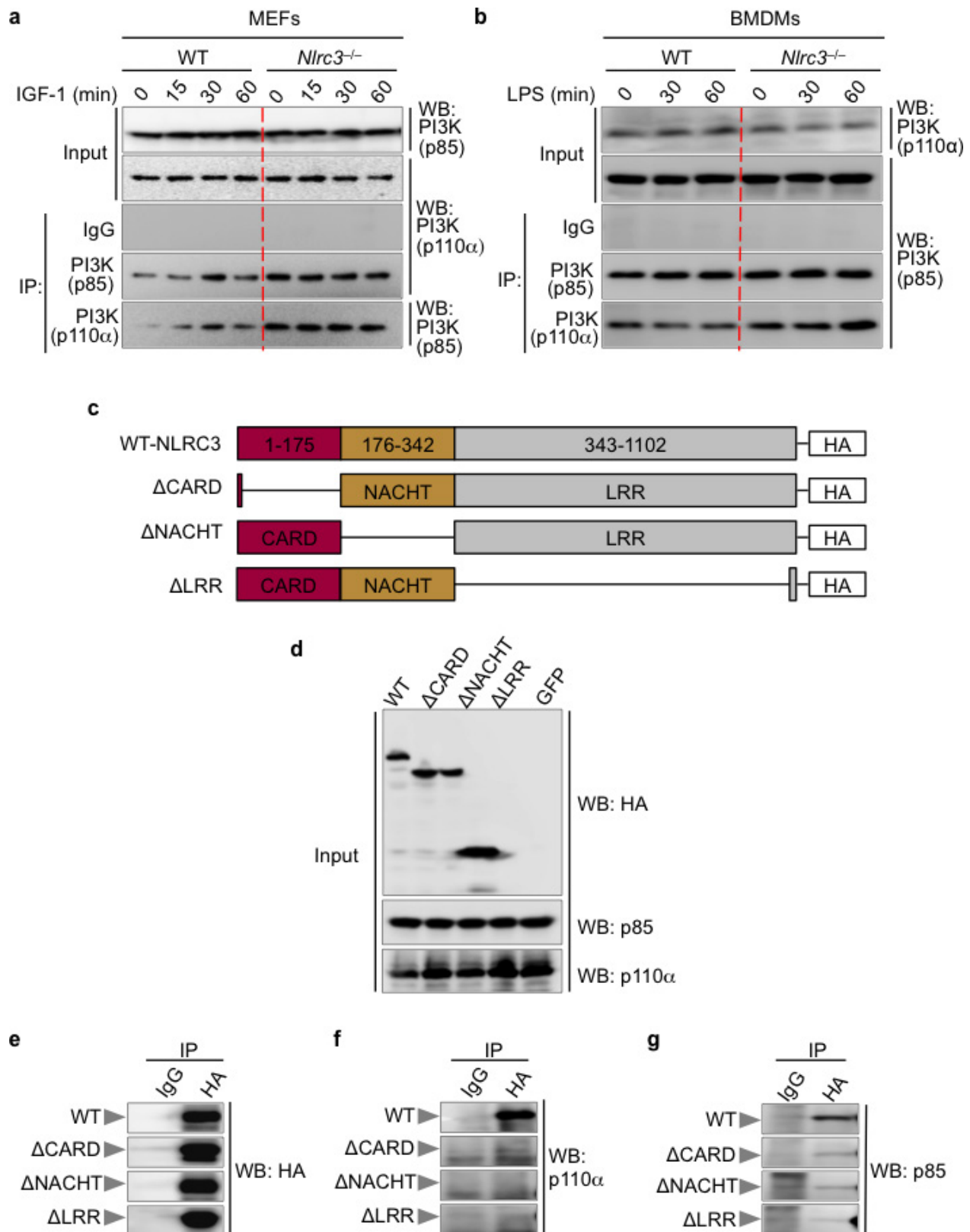
Extended Data Figure 7 | NLRC3 regulates mTOR activity in fibroblasts. **a–d**, Immunofluorescent stain of phosphorylated S6 (**a**), 4E-BP1 (**b**), AKT (**c**), and AKT (**d**) in primary fibroblasts either left untreated or treated with IGF-1 for 30 min is shown in the left panels. Quantification of the fluorescence intensity in each cell ($n = 150$ or more)

is shown in the right panels. ADU, average density unit. Scale bar, 20 μm . Each symbol represents an individual cell. **** $P < 0.0001$; NS, not statistically significant by two-tailed t -test. Data represent one experiment representative of two independent experiments and are presented as mean \pm s.e.m.



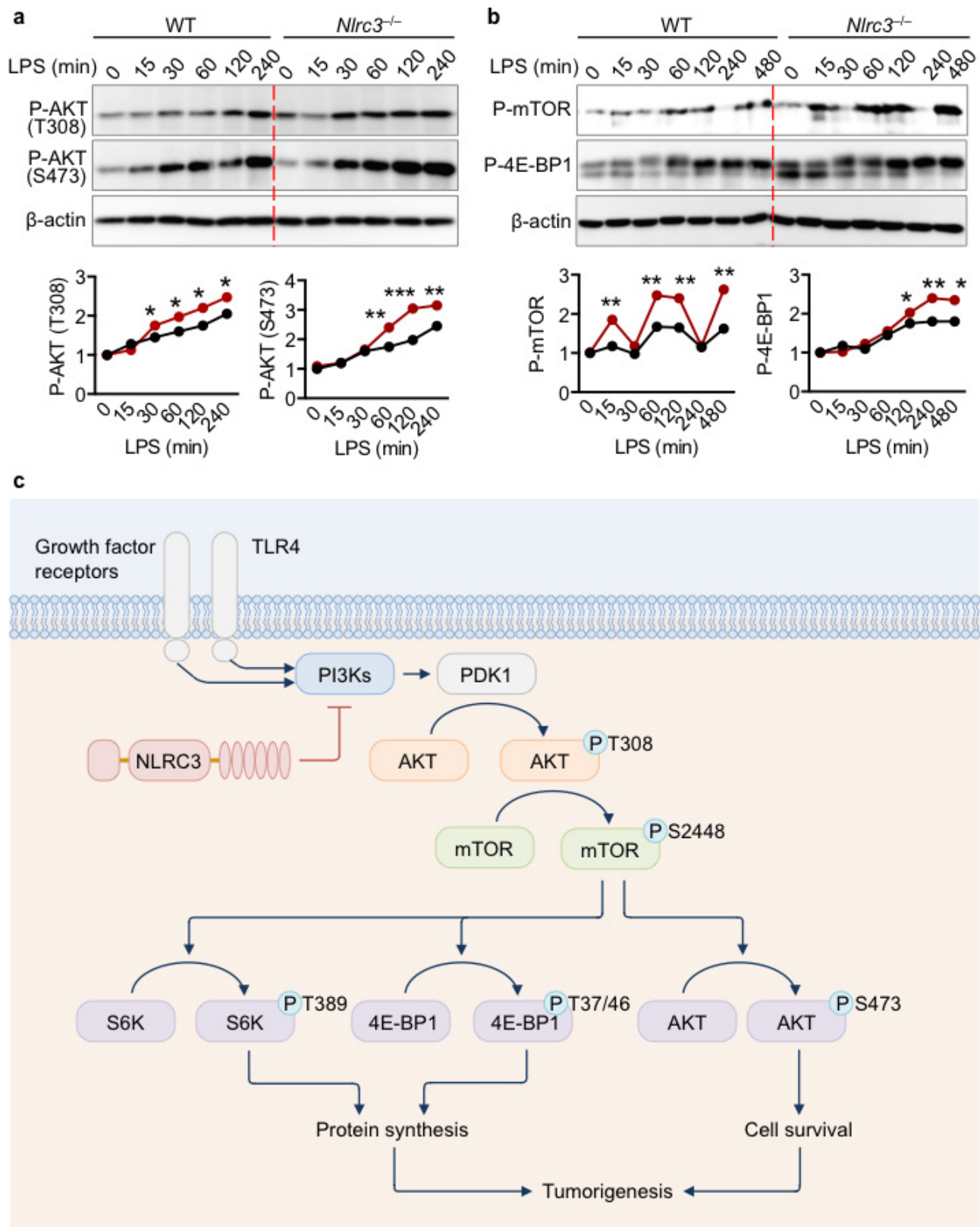
Extended Data Figure 8 | NLRC3 prevents colorectal cancer in an *Apc^{Min/+}* model of tumorigenesis. **a**, Images of colon tumours (left), tumour number and colon length (middle), and size (right) of 120-day-old littermate *Apc^{Min/+}* and *Apc^{Min/+}Nlrc3^{-/-}* mice. **b**, Percentage of mice with dysplasia (left), total histology scores (middle), and histology scores of different parts of colon and different parameters (right) of mice in **a**. **c**, Haematoxylin and eosin (H&E, top), Ki67 (middle) and phosphorylated S6

(P-S6, bottom) staining of colon tumours. Scale bar, 200 μ m. **d**, Quantification of the number (top) and size (bottom) of organoids derived from colonic stem cells of *Apc^{Min/+}* and *Apc^{Min/+}Nlrc3^{-/-}* mice left untreated or treated with LY294002 or rapamycin. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$; NS, not statistically significant by two-tailed t -test (**a**, **b**, **d**). Data represent two independent experiments and are presented as mean \pm s.e.m. in **a**, **b**, **d**.



Extended Data Figure 9 | NLRC3 disrupts the assembly of the PI3K heterodimeric complex. **a, b**, Immunoprecipitation and comparative analysis of the PI3K signalling complex levels between wild-type and *Nlrc3*^{-/-} primary mouse embryonic fibroblasts (MEFs; **a**) and BMDMs (**b**). **c**, Schematic representation of the generation of deletion

mutants of NLRC3. **d**, Loading inputs for **e–g**. **e**, Immunoprecipitation of wild-type-NLRC3 and its deletion mutants. **e, f**, Immunoblotting analysis of the interaction between NLRC3 and its mutants with the PI3K-p110 (**f**) and PI3K-p85 (**g**) subunit. Data represent two independent experiments. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 10 | NLRC3 negatively regulates TLR4-induced activation of the PI3K-AKT-mTOR pathway. **a**, Immunoblot analysis of phosphorylated AKT (at Thr308 and Ser473), total AKT and β -actin (loading control) in wild-type and *Nlrc3*^{-/-} bone-marrow-derived macrophages (BMDMs) left untreated or treated with LPS (top). The protein-band intensity was normalized to β -actin, and expressed relative to that of wild-type controls, set at 1 (bottom). **b**, Immunoblot analysis of phosphorylated mTOR, phosphorylated 4E-BP1, and β -actin

(loading control) in wild-type and *Nlrc3*^{-/-} bone-marrow-derived macrophages (BMDMs) left untreated or treated with LPS (top). Densitometry analysis as in **a** (bottom). **c**, A model of the role of NLRC3 in the negative regulation of the PI3K-AKT-mTOR pathway. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ by two-tailed t -test (**a**, **b**). Data are from one experiment representative of four independent experiments and are presented as mean \pm s.e.m. in **a** and **b**. For gel source data, see Supplementary Fig. 1.

Mechanism of early dissemination and metastasis in Her2⁺ mammary cancer

Kathryn L. Harper^{1*}, Maria Soledad Sosa^{1*†}, David Entenberg², Hedayatollah Hosseini³, Julie F. Cheung¹, Rita Nobre¹, Alvaro Avivar-Valderas¹, Chandandaneep Nagi¹, Nomed Girnius⁴, Roger J. Davis⁴, Eduardo F. Farias¹, John Condeelis², Christoph A. Klein^{3,5} & Julio A. Aguirre-Ghiso¹

Metastasis is the leading cause of cancer-related deaths; metastatic lesions develop from disseminated cancer cells (DCCs) that can remain dormant¹. Metastasis-initiating cells are thought to originate from a subpopulation present in progressed, invasive tumours². However, DCCs detected in patients before the manifestation of breast-cancer metastasis contain fewer genetic abnormalities than primary tumours or than DCCs from patients with metastases^{3–5}. These findings, and those in pancreatic cancer⁶ and melanoma⁷ models, indicate that dissemination might occur during the early stages of tumour evolution^{3,8,9}. However, the mechanisms that might allow early disseminated cancer cells (eDCCs) to complete all steps of metastasis are unknown⁸. Here we show that, in early lesions in mice and before any apparent primary tumour masses are detected, there is a sub-population of Her2⁺p-38^{lo}p-ATF2^{lo}Twist1^{hi}E-cad^{lo} early cancer cells that is invasive and can spread to target organs. Intra-vital imaging and organoid studies of early lesions showed that Her2⁺ eDCC precursors invaded locally, intravasated and lodged in target organs. Her2⁺ eDCCs activated a Wnt-dependent epithelial–mesenchymal transition (EMT)-like dissemination program but without complete loss of the epithelial phenotype, which was reversed by Her2 or Wnt inhibition. Notably, although the majority of eDCCs were Twist1^{hi}E-cad^{lo} and dormant, they eventually initiated metastasis. Our work identifies a mechanism for early dissemination in which Her2 aberrantly activates a program similar to mammary ductal branching that generates eDCCs that are capable of forming metastasis after a dormancy phase.

We investigated whether the loss of tumour-suppressive p38 signalling and gain of oncogenic Her2 upregulation, which induce resistance to anoikis (apoptosis induced by lack of correct cell–extracellular matrix attachment) in early cancer cells¹⁰, might activate a disseminating phenotype. eDCCs were defined as those originating at times when the MMTV–Her2 (Her2 under the control of mouse mammary tumour virus) mice had normal ductal architecture, hyperplasia and mammary intraepithelial neoplasia, as confirmed by histopathology¹¹, but no detectable tumours⁸ (Extended Data Fig. 1a–c). Her2 causes E-cadherin (E-cad, also known as cadherin 1) downregulation¹², whereas p38 can maintain E-cadherin expression¹³. We found that more than 85% of Her2⁺ cells were E-cad^{lo} (Fig. 1a), and in Her2⁺ (wild-type gene) and Her2-T⁺ (mutant active gene) tissues of early lesions, E-cad^{hi} early cancer cells were more frequently phosphorylated (70–75%) (p-) ATF2^{hi} (Fig. 1b and Extended Data Fig. 1d, e). In each duct, 60–70% of all cells from early lesions were positive for membrane β -catenin (the inactive form) (Extended Data Fig. 2a). However, when analysing Her2⁺ cells alone, only 30% of cells showed membrane localization

for β -catenin (Fig. 1c). Overall, these results suggest that Her2⁺ cells display a loss of E-cadherin- and β -catenin-based junctions and are p-ATF2^{lo}.

Overt MMTV–Her2 tumours showed low levels of E-cadherin, phosphorylated p38 (p-p38) and p-ATF2, while maintaining high p-ERK1/2 (p-ERK1 and p-ERK2) levels (Extended Data Fig. 2b–d), suggesting that a Her2⁺p-ATF2^{lo}E-cad^{lo} profile is present in early lesions and primary tumours. We found that only HER2[–] human ductal carcinoma *in situ* (DCIS) lesions retained both high p-ATF2 expression levels and organized E-cadherin junctions, whereas HER2⁺ DCIS lesions showed low expression levels for p-ATF2 and E-cadherin (Extended Data Fig. 2e, f). The Her2⁺p-p38^{lo}p-ATF2^{lo} profile was also present in larger human HER2⁺ breast carcinomas, whereas only HER2[–] tumours showed strong nuclear staining for p-p38 and p-ATF2 compared to HER2⁺ lesions (Extended Data Fig. 2g). We conclude that, in early human and mouse cancer cells, HER2 upregulation is associated with a p-p38^{lo}p-ATF2^{lo}E-cad^{lo} profile that persists in large primary tumours.

The above data suggest that p38 might prevent an invasive phenotype and therefore Her2⁺E-cad^{lo}p-p38^{lo}p-ATF2^{lo} cells from early lesions might be able to disseminate. MMTV–Her2 early lesions and MCF10A (a human mammary gland cell line) cells overexpressing HER2 (denoted as MCF10A-HER2) grown in 3D-organoid cultures show misshapen architecture and occasional single-cell invasion from the organoids (Extended Data Fig. 3a, b). Treatment of MCF10A-HER2 cells (Extended Data Fig. 3c, d) or MMTV–Her2 organoids (Extended Data Fig. 3e) with lapatinib, which reduced p-S6 levels through inhibition of epidermal growth factor receptor (EGFR) and Her2 (Extended Data Fig. 3d), or with HER2 (also known as *ERBB2*) short interfering RNAs (siRNAs) restored E-cadherin junctions and increased p-ATF2 levels (Extended Data Fig. 3c, e). Inhibition of EGFR signalling to AKT using AG1478 also increased the number of E-cadherin junctions and ATF2 phosphorylation (Extended Data Fig. 3c–e). Further, a pan-PI3K inhibitor (GDC-0941) or an AKT inhibitor (MK2266) that reduced p-S6 levels also increased p-ATF2 levels (Extended Data Fig. 3d, e). Basal and lapatinib-stimulated nuclear p-ATF2 levels were completely eliminated by treatment with SB203580 (Extended Data Fig. 3f) in MCF10A-HER2 cells. We conclude that Her2 and EGFR signalling through PI3K and AKT inhibit p38 activity in cells from early lesions, and that increased p-ATF2 levels after HER2 and EGFR inhibition depend on p38 α and p38 β (p38 α / α β) (Extended Data Fig. 3a–f).

MMTV–Her2 and MCF10A-HER2 organoids showed outward invading single cells as observed using live microscopy (Supplementary Video 1 and Fig. 1d, e), which displayed projections rich in F-actin, loss of E-cadherin junctions and focalized loss of laminin-V deposition (Fig. 1e and Extended Data Fig. 3g). Her2⁺ invading cells from

¹Division of Hematology and Oncology, Department of Medicine, Department of Otolaryngology, Department of Oncological Sciences, Tisch Cancer Institute, Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Place, New York, New York 10029, USA. ²Department of Anatomy and Structural Biology, Integrated Imaging Program, Gruss Lipper Biophotonics Center, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, New York, New York 10461, USA. ³Experimental Medicine and Therapy Research, University of Regensburg, 93053 Regensburg, Germany. ⁴Howard Hughes Medical Institute, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. ⁵Project group “Personalized Tumour Therapy”, Fraunhofer Institute for Toxicology and Experimental Medicine, 93053 Regensburg, Germany. [†]Present address: Department of Pharmacological Sciences, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Place, New York, New York 10029, USA.

*These authors contributed equally to this work.

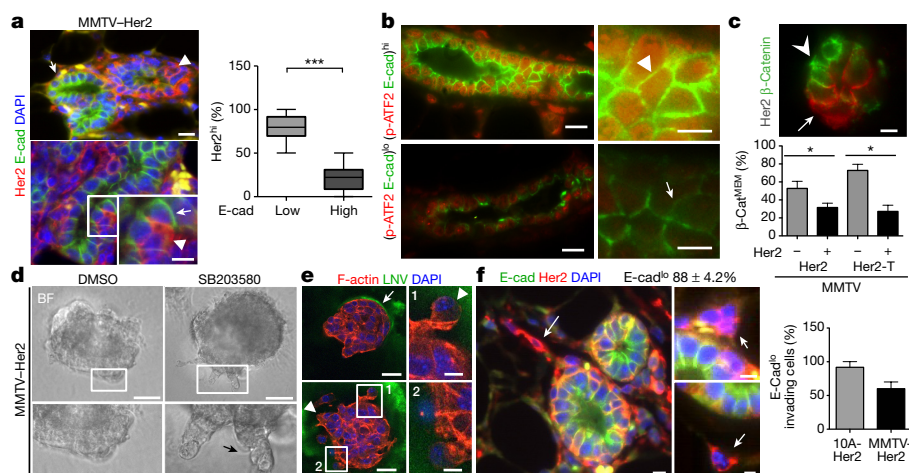


Figure 1 | E-cadherin, Her2 and p-ATF2 levels and function in early lesion cells. **a**, Top, MMTV-Her2 early lesion tissue sections co-stained for the indicated antigens. Bottom, intra-ductal heterogeneity of Her2^{hi}E-cad^{lo} cells. Inset, magnified view of the boxed region. Arrowheads, Her2^{hi}E-cad^{lo} cells; arrows, Her2^{lo}E-cad^{hi} cells. Right, percentage of Her2^{hi} cells that were E-cad^{low} or E-cad^{high} ($n = 20$ ducts; $n = 2$ mice). *** $P < 0.0001$. **b**, Representative images of E-cad^{hi}p-ATF2^{hi} (top) and E-cad^{lo}p-ATF2^{lo} (bottom) ducts in MMTV-Her2 early lesion tissues. Arrowhead, p-ATF2^{hi}E-cad^{hi} cell, arrow, p-ATF2^{lo}E-cad^{lo} cell. Scale bars, 25 μ m and 10 μ m (inset) (**a** and **b**). **c**, MMTV-Her2 early lesion tissue sections stained for Her2 and β -catenin. Arrow, Her2⁺ cells with low membrane-associated β -catenin (β -cat^{MEM-lo}); arrowhead, Her2⁻ β -cat^{MEM-hi} cells. Scale bar, 10 μ m. Graph, percentage of early lesion cells with β -cat^{MEM} that were Her2⁻ or Her2⁺ (Her2, * $P = 0.035$ and

Her2-T, * $P = 0.0008$). MMTV-Her2 $n = 30$ ducts per mouse, $n = 3$ mice; MMTV-Her2-T $n = 10$ ducts per mouse, $n = 2$ mice. **d**, MMTV-Her2 early lesion organoids treated for 48 h with DMSO or 5 μ M SB203580. BF, bright field. Bottom images indicate magnified, boxed regions. Scale bars, 15 μ m (left), 40 μ m (right). **e**, MMTV-Her2 organoids stained for the indicated antigens. Arrowheads, invasive cells; arrow, intact laminin-V layer. Scale bars, 25 μ m (**e**), 10 μ m (insets 1 and 2). **f**, MMTV-Her2 early lesion sections co-stained for the indicated antigens. Arrows, Her2⁺E-cad^{lo} invading cells. Top numbers, percentage of Her2⁺E-cad^{lo} invading cells in early lesion sections, $n = 58$ cells per section, $n = 2$ mice. Graph, percentage E-cad^{lo} invading cells in MCF10A-HER2 (10A-Her2; see also Extended Data Fig. 3a) or MMTV-Her2 organoids. $n = 20$ MCF10A-HER2 organoids. **a**, **c**, one-sided Mann-Whitney U -test. Data are median \pm s.e.m.

MMTV-Her2 early lesions also showed downregulated E-cadherin *in vivo* (Fig. 1f). Notably, approximately 80% of E-cad^{lo} invading cells were Her2⁺ and positive for cytokeratin 8 and 18 (CK8/18⁺), suggesting that epithelial identity was retained despite E-cadherin junction loss; only about 20% of CK8/18⁺ invading cells were negative for Her2 (Extended Data Fig. 3h). When we inhibited p38 α / β activity in MMTV-Her2⁺ and MCF10A-HER2 organoids using SB203580, we found that it stimulated an invasive phenotype (Fig. 1d, e and Extended Data Fig. 3g). We conclude that Her2⁺E-cad^{lo}p-38^{lo} early cancer cells display an invasive phenotype without the loss of CK8/18 expression.

We next monitored early dissemination using high-resolution intra-vital imaging¹⁴ of MMTV-Her2-cyan fluorescent protein (CFP) transgenic mice (Fig. 2). Using a new mammary imaging window and two-photon imaging in vehicle-treated Her2-CFP mice, we found that although no invasion was detected at 10 weeks (normal ductal structure) (Fig. 2a, left and Supplementary Video 2), at 15 and 18 weeks local invasive CFP⁺ cells were detected invading into the stroma (Fig. 2a, middle, right and Supplementary Videos 3, 4). CFP reports faithfully for Her2-overexpressing cells, which were confirmed using double Her2 and CFP immunofluorescence (more than 90% of Her2⁺ cells were CFP⁺; Extended Data Fig. 4a). p38 α / β inhibition with SB203580 stimulated invasion and CFP⁺ cells were now found intravasating (Fig. 2b–d and Supplementary Videos 5–7). 3D reconstruction of the videos showed unambiguously how individual cells from the ductal early lesion tissues entered the lumen of blood vessels (Fig. 2b (inset), c, d and Supplementary Videos 7, 8).

The intravasation documented in the videos led to successful dissemination, since we found Her2⁺CK8/18⁺ eDCCs in the blood, bone marrow and Her2⁺ eDCCs in lungs of 14–18-week-old mice (Extended Data Fig. 4b–j and Extended Data Table 1). Her2 detection by immunohistochemistry or immunofluorescence microscopy using two independent antibodies (mouse and rabbit) showed similar patterns of Her2 staining that were absent with the pre-immune IgG or non-oncogene expressing FVB mice (Extended Data Fig. 4c–g). Systemic inhibition of p38 α / β for 2 weeks also substantially increased

the numbers of early circulating cancer cells in blood (Extended Data Fig. 4h) and eDCCs in bone marrow and lungs (Extended Data Fig. 4i, j).

Treatment of MMTV-Her2 or MCF10A-HER2 early lesion organoids with SB203580 or siRNAs targeting p38 α caused a loss of E-cadherin junctions (Fig. 3a, h (top) and Extended Data Fig. 5a–c). Genetic and pharmacological p38 inhibition reduced total membrane-localized inactive β -catenin and increased active- β -catenin (unphosphorylated β -catenin, detected with a conformation-specific antibody¹⁵) (Fig. 3b and Extended Data Fig. 5a, b, d, e). Inhibition of p38 α / β in MCF10A-HER2 cells led to *AXIN2* mRNA upregulation, a canonical target of β -catenin¹⁶ (Extended Data Fig. 5f). *SNAIL* (also known as *SNAIL*) and *TWIST1* were also upregulated in MCF10A-HER2 3D organoids treated with siRNAs targeting *p38a* (also known as *MAPK14*) or *ATF2* (Extended Data Fig. 5g). Notably, systemic p38 α / β inhibition in MMTV-Her2 mice induced a strong loss of E-cadherin junctions, a concomitant increase of nuclear β -catenin and a strong induction of *Twist1* in early lesion tissues as detected by immunohistochemistry and quantitative PCR (Fig. 3c–f, Extended Data Fig. 5h, i). In wild-type FVB mice or MCF10A organoids treated with or without SB203580 or in wild-type compared to *Mkk3*^{-/-} *Mkk6*^{+/-} (also known as *Map2k3*^{-/-} *Map2k6*^{+/-}) C57BL/6 mice (*Mkk3* and *Mkk6* activate all p38 isoforms), p38 inhibition caused a loss of E-cadherin junctions (Extended Data Fig. 5j–l), but did not cause a loss of membrane β -catenin, CK8/18 expression or α -smooth-muscle-actin-positive myoepithelial cell organization (Extended Data Fig. 5m). Our data suggest that p38 α and ATF2 inhibit the activity of β -catenin, preventing successful dissemination of Her2⁺ early lesion cells. Our data also suggest that in the absence of Her2 expression, p38 inhibition was insufficient to change β -catenin localization.

HER2 expression or p38 α / β inhibition alone induced a 14-gene EMT signature that included non-canonical WNT ligands and EMT transcription factors and was further upregulated by p38 inhibition in MCF10A-HER2 cells (Fig. 3g and Extended Data Fig. 6a, b). *CDH1* mRNA was also downregulated by p38 inhibition in MCF10A-HER2

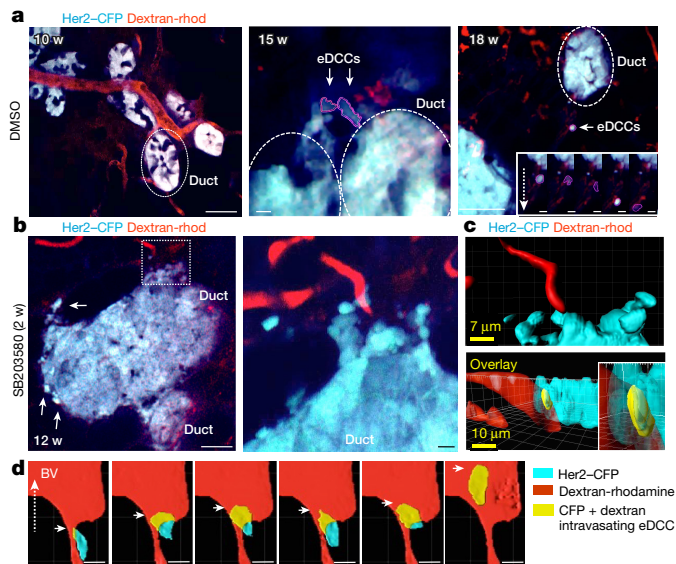


Figure 2 | Intra-vital imaging of eDCC precursor intravasation. **a**, Her2-CFP (blue) ducts from 10- (left, Supplementary Video 2), 15- (middle, Supplementary Video 3) and 18- (right, Supplementary Video 4) week-old early lesions; scale bars, 38, 5 and 56 μm , respectively. Red, rhodamine-dextran-labelled vasculature. Dotted ellipses, ducts. Right inset: direction of movement of an eDCC at 5 time points (2 min per panel). Dotted arrow, direction of movement; scale bars in insets, 7.8 μm . **b**, Her2-CFP early lesion ducts from Supplementary Video 5 and Supplementary Video 6 (inset) treated with SB203580 for 2 weeks. Middle image, Higher resolution intra-vital video of the area boxed in the left image. Scale bar, 25.5 μm and 5.2 μm (boxed area). **c**, Top, 3D computer reconstruction of the video in **b** (boxed area). Bottom, rotated projection showing invasion (yellow) of early lesion cells (CFP, cyan) into blood vessels (red) (Supplementary Video 7). Scale bars, 7 μm (top) and 10 μm (bottom). **d**, Sequence following an early lesion cell (Supplementary Video 8) as it intravasates in mice treated with SB203580 for 4 weeks (yellow cell inside the red blood vessel (BV); 2.2 min per panel).

organoids (Extended Data Fig. 6c). When we determined whether WNT ligands were functionally linked to the HER2-driven EMT-like program, we found that MCF10A-HER2 cells treated with SB203580 displayed *AXIN2* mRNA induction, which was reversed by overexpression of the WNT-ligand antagonist SFRP1 (Extended Data Fig. 6d, e). Recombinant soluble WNT3A (ref. 17 and Methods) also stimulated expression of *AXIN2* in MCF10A-HER2 cells, and this induction was substantially inhibited by expression of a constitutively active p38 α kinase (p38 α ^{D176A,F327S} mutant) cDNA¹⁰ (Extended Data Fig. 6f). Increased invasion, loss of E-cadherin junctions and membrane-localized β -catenin after p38 α / β inhibition were reversed in SFRP1-expressing cells (Extended Data Fig. 6g, h). In addition, the canonical Wnt inhibitor DKK1 also reversed the loss of E-cadherin induced by p38 inhibition in MMTV-Her2 primary organoids (Fig. 3h). *Twist1* was also readily detected as a nuclear-cluster signal using immunofluorescence microscopy in both early lesions and primary tissue from primary tumours in the MMTV-Her2 model (Extended Data Fig. 6i). Analysis of these tissues showed that *Twist1* was expressed in the majority of early lesion cells and that no major changes in the expression of *Twist1* occurred with progression to primary tumours. Overall, our data suggest that Her2 and p38 operate antagonistically and that Her2⁺p-p38^{lo} cells from early lesions may rely on both canonical and non-canonical Wnt ligands to induce an EMT-like program associated with TWIST1 upregulation.

The majority of solitary Her2⁺ eDCCs (1–5 cells) in lungs were negative or weakly stained for the G1 exit marker p-Rb (p-retinoblastoma protein) compared to growing micro and macro-metastases in mice carrying primary tumours (Fig. 4a, b, d). Detection of p-Ser10 histone-H3 (a G2/M marker) also showed that most eDCCs were

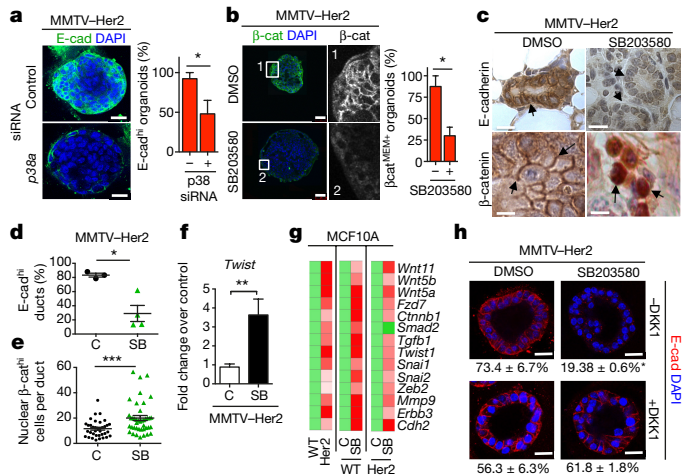


Figure 3 | An EMT-like program in Her2⁺ early lesion cells.

a, E-cadherin staining in early lesion MMTV-Her2 organoids transfected with control or p38 α -targeting siRNA. Scale bars, 25 μm . Graph, percentage E-cad^{hi} organoids; $n = 30$ organoids per treatment; * $P = 0.038$. **b**, Early lesion MMTV-Her2 organoids treated for 48 h with SB203580 and stained for total β -catenin and the nuclei were counterstained with DAPI. Grey-scale panels (1, 2) denote zoomed images of boxed areas showing membrane- and cytosolic-localized β -catenin. Scale bars, 50 μm . Graph, percentage β -cat^{MEM+} organoids; $n = 10$ organoids per treatment; * $P = 0.034$. **c**, Early lesion tissue sections from MMTV-Her2 mice treated for 2 weeks with SB203580 stained for E-cadherin and β -catenin. Scale bars, 20 μm (top) and 10 μm (bottom). Arrows, E-cadherin (top left) and lack of signal (top right), membrane-bound β -catenin (bottom, left, arrows) or nuclear β -catenin (bottom, right, arrows). **d**, **e**, Quantification of images in **c**. C, control. SB, SB203580. **d**, Percentage of E-cad^{hi} ducts. $n = 3$ animals; * $P = 0.028$. **e**, Percentage of nuclear β -cat^{hi} cells per duct. $n = 36$ ducts per 3 animals; *** $P = 0.0002$. **f**, *Twist1* mRNA levels in MMTV-Her2 early lesion tissues obtained from mice treated as in **c**. Values, fold change over control (DMSO) normalized to *Gapdh*. $n = 3$ mice per treatment; ** $P = 0.0042$. **g**, Heat map of EMT-related genes upregulated >2-fold (biological triplicate) in MCF10A and MCF10A-HER2 organoids treated for 6 days with or without SB203580 (5 μM). Green, control values set to 1; red, fold change over control. **h**, Early lesion MMTV-Her2 organoids were treated for 2 days with SB203580 (5 μM) as well as 500 ng ml⁻¹ DKK1 and stained for E-cadherin. Bottom numbers, percentage of E-cad^{hi} organoids; $n = 10$ organoids per treatment, biological duplicates. * $P < 0.01$ (DMSO – DKK versus SB203580 – DKK and SB203580 – DKK1 versus SB203580 + DKK1); not significant (DMSO – DKK1 versus DMSO + DKK1; DMSO + DKK1 versus + DKK1 and SB203580). Scale bars, 25 μm . **a**, **b**, **f**, **h**, One-sided unpaired *t*-test; **d**, **e**, One-sided Mann–Whitney *U*-test. Data are mean \pm s.e.m.

non-proliferative (Fig. 4d and Extended Data Fig. 6j). DCCs found in animals bearing overt tumours were termed DCCs because we could not distinguish their early lesion or primary tumour origin. Animals with overt tumours that were p-Rb^{strong} (Extended Data Fig. 6k) and lungs bearing proliferative (p-Rb^{strong}) micro- (6–50 Her2⁺ cells) and macro- (more than 50 Her2⁺ cells) metastases (Fig. 4b) still had numerous quiescent p-Rb[–]Her2⁺ DCCs (Fig. 4a, d). We conclude that eDCCs are primarily p-Rb[–]p-H3[–] and that even in animals bearing metastases, more than 60% of single or less than 5-cell clusters are non-proliferative. Notably, the vast majority of solitary Her2⁺ eDCCs in lungs were negative for E-cadherin (Fig. 4e). However, Her2⁺ DCCs (in animals with overt primary tumours) were positive for E-cadherin in approximately 48% of the population (Fig. 4e). Also, close to 100% of the solitary Her2⁺ eDCCs showed high *Twist1* expression, whereas only 30% of solitary Her2⁺ DCCs in animals with primary tumours were positive for *Twist1* (Fig. 4f and Extended Data Fig. 6l). We conclude that most Her2⁺ eDCCs are quiescent and upregulate *Twist1*, whereas more than half of the DCCs during primary tumour stages downregulate *Twist1* expression, suggesting that reactivation might be linked to a mesenchymal–epithelial transition¹⁸.

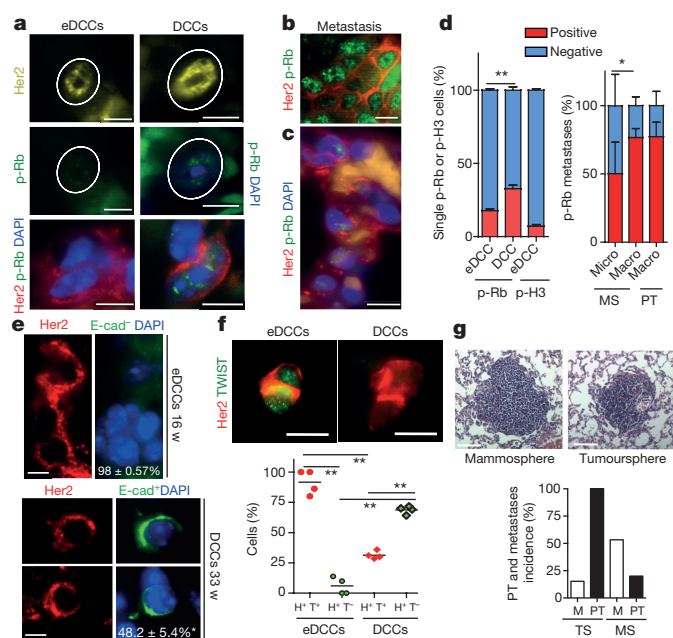


Figure 4 | eDCC characterization and metastatic potential. **a**, Her2, p-Rb and DAPI immunofluorescence detection in eDCCs and DCCs in lung sections from MMTV-Her2 mice. Scale bars, 10 μ m. **b**, Her2 and p-Rb levels in spontaneous metastasis in MMTV-Her2 mice carrying autochthonous primary tumours. Scale bars, 10 μ m. **c**, Her2, p-Rb and DAPI detection in micro-metastasis generated from early lesion mammospheres. Scale bars, 10 μ m. **d**, Left, percentage of positive or negative p-Rb or p-Ser10 histone 3(H3) and Her2⁺ solitary eDCCs or DCCs; $n = 3$ animals per group; mean \pm s.e.m.; ** $P = 0.0087$; one-sided, unpaired t -test. Right, Percentage of positive or negative p-Rb cells within spontaneous macro-metastases ($n = 3$ lesions) in Her2 mice carrying overt autochthonous primary tumours (PT) or in metastases derived from MMTV-Her2 early lesions mammospheres (MS) ($n = 9$ micro-, $n = 3$ macro-metastasis). * $P = 0.021$, one-sided, unpaired t -test; mean \pm s.d. from technical replicates. **e**, Her2, E-cadherin and DAPI signal in eDCCs or DCCs; bottom numbers, percentage of E-cad⁺ (top) or E-cad⁺ (bottom) DCCs, $n \geq 100$ DCCs from 3 mice. Scale bars, 5 μ m. **f**, Left, Her2⁺ (H⁺) and Twist1⁺ (T⁺) eDCCs; right, Her2⁺Twist1⁻ (T⁻) DCCs. Graph, percentage of cells with the indicated profiles. $n = 500$ cells; 4 animals per group; ** $P < 0.05$; one-sided Mann-Whitney U -test; median and values from individual animals. Scale bars, 10 μ m. **g**, Haematoxylin and eosin staining of lung macro-metastasis produced by MMTV-Her2 mammospheres (MS) or tumourspheres (TS). Scale bars, 150 μ m. Graph, primary tumour (PT) and metastasis (M) percentage incidence of early lesions, mammospheres and primary tumourspheres. $n = 15$ (MS) and 13 (TS) animals.

The treatment of MMTV-HER2 mice with p38 inhibitors during early lesion stages for 2 weeks did not stimulate the dormant eDCCs to proliferate in the lungs as measured by p-Rb staining (Extended Data Fig. 6m). However, if p38 was inhibited for 2 weeks during overt primary tumour stages, DCCs responded to the p38 inhibitor and the proportion of p-Rb⁺ cells was increased (Extended Data Fig. 6m). These data suggest that, in agreement with Twist1 upregulation in response to p38 inhibition, the non-proliferative state of Twist1^{hi} eDCCs is independent of p38 α / β activation. However, DCCs in animals with overt tumours regain sensitivity to p38 α / β inhibition as observed in other models¹⁹.

We next analysed the tumour- and metastasis-initiating capacity of pure cancer cells from early lesions. We prepared sphere cultures from MMTV-Her2 early lesions or from overt primary tumour tissues (Extended Data Fig. 7a) and found that early lesion mammospheres were more efficient at generating secondary spheres than primary tumoursphere cells (Extended Data Fig. 7b). However, after orthotopic injection into nude mouse mammary fat pads (approximately

300 spheres per site), 100% of primary tumourspheres formed primary tumours within 4–12 weeks (Fig. 4g), whereas Her2⁺ early lesion mammospheres did not produce obvious growing tumours (Fig. 4g and Extended Data Fig. 8); in three animals Her2⁺ early lesion mammospheres formed small nodules at two months but these entered stasis. The tumorigenic capacity of primary tumours over early lesion spheres correlated with enhanced p-ERK1/2 and p-S6 levels in the former (Extended Data Figs 7d, 8). Notably, Her2⁺ early lesion mammospheres produced lung metastases that were Her2⁺ and p-Rb^{strong}, confirming their early lesion MMTV-Her2 origin and that they can proliferate (Fig. 4c, d, g and Extended Data Fig. 7c). Spontaneous macro-metastases in MMTV-Her2 mice bearing autochthonous tumours were Her2⁺ and p-Rb⁺ (Fig. 4b, d) and macro-metastases derived from MMTV-Her2 early lesion mammospheres were also positive for p-Rb (Fig. 4c, d). Her2⁺ tumourspheres produced Her2⁺ DCCs (Extended Data Fig. 7c) and also showed macro-metastasis but with a lower incidence (around 16%) (Fig. 4g).

Mammospheres, which were re-implanted in 3D matrices and subsequently imaged, were substantially more invasive than tumourspheres that remained globular (Extended Data Fig. 7e). We also found that experimental metastasis incidence from single-cell suspensions was 100% in both groups and no significant difference was found in the ability of early lesion or primary tumour cancer cells to produce metastatic nodules (Extended Data Fig. 7f). Our data suggest that Her2⁺ early lesion cells that downregulate p38 signalling and activate an EMT-like response while largely non-tumorigenic in orthotopic sites, successfully activate invasive programs that allow for efficient dissemination and metastasis formation. The latter does not seem to only depend on the ability to colonize lungs (Fig. 4c, g and Extended Data Fig. 7c, f).

Our findings identify a molecular mechanism of early dissemination. We propose that a subpopulation of Her2⁺CK8/18⁺Wnt^{hi}p-p38^{lo}Twist1^{hi}E-cad^{lo} early cancer cells can disseminate and metastasize (Extended Data Fig. 9). Early lesion cells are more invasive than primary-tumour-derived cells, display more stemness and can intravasate and lodge in secondary organs. Notably, in addition to the work presented here, Hosseini *et al.*³¹ also show that eDCCs have metastasis-initiating capacity, which was not associated with high tumorigenic capacity or enhanced lung-colonizing potential. Analysis of ERK1/2 and AKT-mTOR pathways showed low activation of these pathways in early lesion cancer cells, suggesting they are not in a ‘growth mode’ yet. This suggests a new function for Her2, in which, prior to stimulating growth in early lesions, it activates programs of stemness, motility and invasion similar to those observed during branching morphogenesis²⁰ that support early dissemination and metastasis.

A few studies have shown that early dissemination occurs in mammary-cancer⁸, pancreatic-cancer⁶ and melanoma⁷ models and in patients^{3,21}, suggesting that our findings are not a rarity of the Her2 model. The majority (around 98%) of Her2⁺ single eDCCs, once in the lungs, were Her2⁺Twist1^{hi} but E-cad^{lo} and predominantly dormant. This suggests that eDCC dormancy might be linked to an EMT program, as proposed by others^{22,23}. However, eDCCs still expressed CK8/18⁺, suggesting that a partial EMT is sufficient for early dissemination, dormancy and metastasis formation. This EMT program in eDCCs may allow cells to interconvert between dormant and proliferative states as transient Twist1 expression induces stem-cell programs²⁴ while a full EMT may block metastasis^{24,25}. Whether changes in Twist1 and E-cadherin expression control the reactivation in lungs remains unknown.

Our work also reveals an unexpected role for p38 α / β kinases²⁶ and ATF2 (ref. 27) in antagonizing Her2 signalling early in cancer progression. E-cadherin junction regulation by p38-mediated¹³ and/or ATF2-mediated blockade of β -catenin activity²⁸, may explain how these proteins block early dissemination. eDCC precursors showed low p38 activation and eDCCs were not stimulated to proliferate after systemic p38 α / β inhibition, which stimulated expansion of DCCs in other models¹. Thus, eDCC dormancy seems to be p38 α / β independent.

By contrast, p38 appeared to regulate DCC dormancy during late stages of progression as shown in other models¹. The mechanism behind this p38 switch between eDCCs and DCCs remains unknown. The data in the study by Hosseini *et al.*³¹, using genetic lineage analysis, showed that a large proportion of metastases are derived from eDCC ancestors. Thus, understanding the differences between eDCCs and DCCs is important to better adapt the targeting of these different DCC populations.

Our findings change our understanding of how certain oncogenes may initiate dissemination before triggering aggressive proliferation and how tumour-suppressor pathways might suppress metastasis, a function that has been attributed to p38 (ref. 1), but was never linked to the early dissemination process. The related study³¹ and our work may also open doors to explain phenomena like metastases in cancer of unknown primary origin²⁹ and in patients with DCIS that never developed any local recurrence³⁰. We may also be able to understand how eDCCs found metastasis directly and/or through the preparation of eDCC-mediated pre-metastatic niches for later arriving DCCs to colonize target organs. These findings might inform on better ways to target DCCs in all their forms to prevent metastasis.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 October 2015; accepted 3 November 2016.

Published online 14 December 2016.

- Sosa, M. S., Bragado, P. & Aguirre-Ghiso, J. A. Mechanisms of disseminated cancer cell dormancy: an awakening field. *Nat. Rev. Cancer* **14**, 611–622 (2014).
- Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
- Schardt, J. A. *et al.* Genomic analysis of single cytokeratin-positive cells from bone marrow reveals early mutational events in breast cancer. *Cancer Cell* **8**, 227–239 (2005).
- Klein, C. A. *et al.* Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc. Natl Acad. Sci. USA* **96**, 4494–4499 (1999).
- Schmidt-Kittler, O. *et al.* From latent disseminated cells to overt metastasis: genetic analysis of systemic breast cancer progression. *Proc. Natl Acad. Sci. USA* **100**, 7737–7742 (2003).
- Rhim, A. D. *et al.* EMT and dissemination precede pancreatic tumor formation. *Cell* **148**, 349–361 (2012).
- Eyles, J. *et al.* Tumor cells disseminate early, but immunosurveillance limits metastatic outgrowth, in a mouse model of melanoma. *J. Clin. Invest.* **120**, 2030–2039 (2010).
- Hüsemann, Y. *et al.* Systemic spread is an early step in breast cancer. *Cancer Cell* **13**, 58–68 (2008).
- Sänger, N. *et al.* Disseminated tumor cells in the bone marrow of patients with ductal carcinoma *in situ*. *Int. J. Cancer* **129**, 2522–2526 (2011).
- Wen, H. C. *et al.* p38 α signaling induces anoikis and lumen formation during mammary morphogenesis. *Sci. Signal.* **4**, ra34 (2011).
- Cardiff, R. D. Validity of mouse mammary tumour models for human breast cancer: comparative pathology. *Microsc. Res. Tech.* **52**, 224–230 (2001).
- Lu, J. *et al.* 14-3-3 ζ cooperates with ErbB2 to promote ductal carcinoma *in situ* progression to invasive breast cancer by inducing epithelial–mesenchymal transition. *Cancer Cell* **16**, 195–207 (2009).
- Strippoli, R. *et al.* p38 maintains E-cadherin expression by modulating TAK1–NF- κ B during epithelial-to-mesenchymal transition. *J. Cell Sci.* **123**, 4321–4331 (2010).
- Entenberg, D. *et al.* Setup and use of a two-laser multiphoton microscope for multichannel intravital fluorescence imaging. *Nat. Protocols* **6**, 1500–1520 (2011).
- Malladi, S. *et al.* Metastatic latency and immune evasion through autocrine inhibition of WNT. *Cell* **165**, 45–60 (2016).

- Leung, J. Y. *et al.* Activation of AXIN2 expression by β -catenin–T cell factor. A feedback repressor pathway regulating Wnt signaling. *J. Biol. Chem.* **277**, 21657–21665 (2002).
- Grumolato, L. *et al.* Canonical and noncanonical Wnts use a common mechanism to activate completely unrelated coreceptors. *Genes Dev.* **24**, 2517–2530 (2010).
- Nieto, M. A., Huang, R. Y., Jackson, R. A. & Thiery, J. P. EMT: 2016. *Cell* **166**, 21–45 (2016).
- Bragado, P. *et al.* TGF- β 2 dictates disseminated tumour cell fate in target organs through TGF- β -RIII and p38 α / β signalling. *Nat. Cell Biol.* **15**, 1351–1361 (2013).
- Brisken, C. *et al.* Essential function of Wnt-4 in mammary gland development downstream of progesterone signaling. *Genes Dev.* **14**, 650–654 (2000).
- Rhim, A. D. *et al.* Detection of circulating pancreas epithelial cells in patients with pancreatic cystic lesions. *Gastroenterology* **146**, 647–651 (2014).
- Ocaña, O. H. *et al.* Metastatic colonization requires the repression of the epithelial–mesenchymal transition inducer Prrx1. *Cancer Cell* **22**, 709–724 (2012).
- Brabletz, T. To differentiate or not – routes towards metastasis. *Nat. Rev. Cancer* **12**, 425–436 (2012).
- Schmidt, J. M. *et al.* Stem-cell-like properties and epithelial plasticity arise as stable traits after transient Twist1 activation. *Cell Reports* **10**, 131–139 (2015).
- Fischer, K. R. *et al.* Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472–476 (2015).
- Hui, L. *et al.* p38 α suppresses normal and cancer cell proliferation by antagonizing the JNK–c-Jun pathway. *Nat. Genet.* **39**, 741–749 (2007).
- Gozdecka, M. *et al.* JNK suppresses tumor formation via a gene-expression program mediated by ATF2. *Cell Reports* **9**, 1361–1374 (2014).
- Bhoumik, A. *et al.* Suppressor role of activating transcription factor 2 (ATF2) in skin cancer. *Proc. Natl Acad. Sci. USA* **105**, 1674–1679 (2008).
- Fizazi, K. *et al.* Cancers of unknown primary site: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Annals Oncol.* **26**, v133–v138 (2015).
- Narod, S. A., Iqbal, J., Giannakeas, V., Sopik, V. & Sun, P. Breast cancer mortality after a diagnosis of ductal carcinoma *in situ*. *JAMA Oncol.* **1**, 888–896 (2015).
- Hosseini, H. *et al.* Early dissemination seeds metastasis in breast cancer. *Nature* <http://dx.doi.org/10.1038/nature20785> (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Parsons and P. Polulidakos for PI3K and AKT inhibitors, S. Aaronson and H.-C. Wen for WNT3A, SFRP1 and DKK1 reagents and expertise. Grant support: HHMI (R.J.D.), SWCRF (J.A.A.-G. and E.F.F.), CA109182, CA196521 (J.A.A.-G.), CA163131 (J.A.A.-G. and J.C.), CA100324 (J.C.), F31CA183185 (K.H.), BC132674 (J.A.A.-G. and J.C.), BC112380 (M.S.S.), NIH 1S10RR024745. Microscopy CoRE at ISMMS. DFG KL 1233/10-1 and the ERC (322602) (C.A.K.).

Author Contributions K.L.H. designed, performed experiments, analysed data and co-wrote the manuscript; M.S.S. designed experimental approach, performed experiments, executed intravital imaging, provided oversight, analysed data and co-wrote the manuscript; D.E. designed and executed intravital imaging, analysed data and co-wrote the manuscript; H.H. provided materials and analysed data; A.A.V. performed experiments; C.N. provided materials and histopathological analysis; J.F.C. managed mouse colonies and performed experiments; R.N. performed experiments and analysed data; N.G. maintained the *Mkk3/Mkk6* wild-type and knockout mice and provided materials; R.J.D. provided materials and co-wrote manuscript; C.A.K. provided input for the writing of the manuscript; J.C. designed intra-vital experiments, analysed data and co-wrote the manuscript; E.F.F. provided expertise and analysed data; J.A.A.-G. designed and optimized experimental approach, provided general oversight, collected microscopy data, analysed data and co-wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S.S. (maria.sosa@mssm.edu) or J.A.A.-G. (julio.aguirre-ghiso@mssm.edu).

Reviewer Information *Nature* thanks M. Bissell, C. Ghajar and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Cells and cell culture. MCF10A cells were obtained from J. Brugge and ATCC (MCF10A; ATCC CRL-10317) for verification of phenotype. MCF10A cells were authenticated by cultures in 3D matrigel as reported previously¹⁰. MCF10A-HER2 cells were transfected with HER2 plasmids obtained from L. Petty and the identity of the vectors was confirmed by sequencing. HER2-expressing cells were selected for by the addition of G418. MCF10A cells expressing SFRP1 were generated using SFRP1 lentiviral vectors. WNT3A- and DKK1-conditioned media was prepared from WNT3A-expressing L-cells and DKK1-expressing 293T cells. Vectors and SFRP1 plasmids were a gift from S. Aaronson. Conditioned media was prepared from cells cultured with serum-free medium (DMEM with 1% penicillin and streptomycin) for 24 h and then concentrated using Vivaspin 20 Centrifugal Concentrating tubes (Sartorius, VS2021) at 3,000g for up to 3 h until the desired concentration ($10\times$) was reached.

Mammospheres and tumourspheres assays. Animal procedures were approved by the Institutional Animal Care and Use Committee (IACUC) of Icahn School of Medicine at Mount Sinai, protocol 08-0366. MMTV-Her2 mice were euthanized using CO₂ at 14–18 weeks of age or when overt tumours had formed (primary tumours). For mammosphere preparations, all 5 pairs of glands in mice were checked for the presence of any visible small lesions or palpable tumours when processed for early cancer cells and none were found. Even when other mammary glands were inspected microscopically in whole mounts we could not detect small tumours. Whole mammary glands or tumours were digested in collagenase and bovine serum albumin (BSA) at 37°C for 45–60 min. Red-blood-cell lysis buffer was used to remove blood cells from cultures and cells were then plated for 10–15 min in DMEM containing 10% fetal bovine serum in 35-mm dishes at 37°C for fibroblast removal. Cells were then incubated in 2 mM PBS-EDTA for 15 min at 37°C and passed through a 25-gauge needle. Cell suspensions were then filtered through a 70- μ m filter before counting. Cells were seeded in 6-well ultra-low-adhesion plates at a density $>5.0 \times 10^5$ cells per well in 1 ml mammosphere media (DMEM/F12 (Gibco 11320-082), 1:50 B27 (Invitrogen 17504-044), 20 ng ml⁻¹ EGF (Peprotech AF-100-15-A), 1:100 penicillin and streptomycin).

Animal experiments. Animal procedures were approved by the Institutional Animal Care and Use Committee (IACUC) of Icahn School of Medicine at Mount Sinai protocol 08-0366. Tumours were not allowed to grow beyond the IACUC allowed limit of 1,000 mm³ per animal. Animals were randomized and assigned to a group when they reached a certain age, so there is no prior knowledge or assumption when assigning the mice to treatments. Approximately 300 spheres from early lesions or primary tumours were injected per site into nude mice (BALB/c^{nu/nu}, Charles River). Suspension cultures were spun at 300 r.p.m. for 4 min and then suspended in 150 μ l PBS Ca²⁺ and Mg²⁺ per 300 spheres. Matrigel (Corning 356231) was then added in a 1:1 ratio. Spheres were injected into the two fourth inguinal gland fat pad using a 27-gauge needle. In the case of mice injected with tumour-derived spheres, mice were euthanized when the tumour reached 1,000 mm³ according to IACUC regulations. Incidence measured at 1, 3 and 12 months was calculated. The 12-month time point was not assessed for the tumour-sphere group. Micro-metastases, 3–20 cell clusters. Macro-metastases, clusters with more than 20 cells per cluster.

3D mammary primary epithelial cell and MCF-10A organoid cultures. MMTV-Her2 mice were euthanized using CO₂ at 14–18 weeks of age and mammary epithelial cells (MECs) were isolated using the same protocol used for mammosphere preparation. Similar to mammosphere preparations, all 5 pairs of glands in mice were checked for the presence of any visible or palpable tumours when processed for early cancer cells and none were found, even in whole mounts prepared from mammary glands from the same mouse processed for mammosphere preparation. 5.0×10^4 MECs were seeded in 400 μ l assay medium (DMEM/F12, 5% horse serum, 1% penicillin and streptomycin, and 20 ng ml⁻¹ EGF plus 2% Matrigel) in 8-well chamber slides with 40 μ l of Matrigel. Organoids formed at an efficiency of around 30 organoids per 1.0×10^4 MECs plated. MCF10A-HER2 cells were cultured in three-dimensional cultures as previously described^{32,33}. In brief, cells were seeded in 400 μ l assay medium consisting of DMEM/F12, 5% horse serum, 1% penicillin and streptomycin, and EGF plus 2% Matrigel on 40 μ l Matrigel (Corning) in 8-well chamber slides (Falcon 354108).

Treatments of organoids cultures. Cultures were treated every 24 h starting at day 6 with 5 μ M SB203580 (Calbiochem, 559395). 6×10^3 cells were seeded for immunofluorescence studies and were fixed at day 12 with 4% paraformaldehyde (PFA) or 10% formalin with phosphatase and protease inhibitors (NaVO₃, NaF, pepstatin A, leupeptin and aprotinin). To measure mRNA changes in 3D cultures, 5.0×10^4 cells were seeded in 1 ml Matrigel in a 24-well plate and RNA was extracted from cultures at day 12 using 1 ml Trizol (Ambion 15596018) followed by RNA extraction. Cultures were treated every 24 h with 5 μ M DMSO or SB203580 and 500 ng ml⁻¹ DKK1 for 48 h following organoids formation and

fixed for immunofluorescence with 4% PFA or 10% formalin with phosphatase and protease inhibitors. In some cases spheres growing in suspension conditions were seeded in assay medium plus 2% of Matrigel directly in 40 μ l of Matrigel and 3 days later the number of invasive spheres was counted.

Differential interference contrast microscopy. 3D matrigel organoids were time-lapse imaged using a Zeiss LSM 880 Confocal with Airyscan microscope $10\times/0.45$ objective lens with $4\times$ zoom. We recorded four positions per condition in parallel for 2 h with 20 min interval varying with the depth of the cell. Temperature was maintained at 37°C and CO₂ at 5%. Zen 2.1 software was used to acquire and export images as an uncompressed AVI. Videos were made by using either Zen 2.1 or ImageJ software. This was done using the services of the Microscopy Facility at Icahn School of Medicine at Mount Sinai.

Immunofluorescence. 3D cultures were fixed with 4% PFA for 20 min at room temperature in the presence of phosphatase and protease inhibitors. Staining was performed as previously described for MCF10A 3D cultures³². In brief, cells were permeabilized using 0.1% Triton X-100 in PBS for 20 min. Blocking was done using $1\times$ immunofluorescence PBS wash buffer (130 mM NaCl; 7 mM Na₂HPO₄; 3.5 mM NaH₂PO₄; 7.7 mM NaN₃; 0.1% BSA; 0.2% Triton X-100; 0.05% Tween-20) containing 10% normal goat serum (Gibco, PCN5000) for 1 h. Primary antibodies used were: E-cadherin (BD Biosciences 610181); β -catenin (BD Biosciences 610153); laminin V (Progen 10765); F-Actin (Life Technologies A12380); α -smooth muscle actin (Sigma-Aldrich C6198); active β -catenin (Millipore 05-665); CK8/18 (ProGen GP11). The following secondary antibodies were used: AlexaFluor goat-anti-mouse 488, AlexaFluor goat-anti-rabbit 568. Chambers were removed from slides and wells were fixed and mounted with ProLong Gold Antifade reagent with DAPI (Invitrogen P36931). 2D cultures were fixed with 4% PFA for 20 min at 4°C. Cells were permeabilized in 0.1% TritonX-100 and then blocked in 3% normal goat serum for 30 min at room temperature. Primary antibodies were left 1 h at room temperature in 0.1% BSA in PBS followed by an additional blocking step. Secondary antibodies were left for 1 h at room temperature in 0.1% BSA in PBS. Cover slips were fixed using ProLong Antifade mounting media with DAPI (Molecular Probes P36930). Primary antibodies used were: E-cadherin (BD Biosciences, 610181) and p-ATF2 (Cell Signaling, 9226). Imaging of 3D organoids was done using confocal microscopy. Images were obtained using Leica Software on a Leica SP5 confocal microscope. Mammary gland section imaging was done using a Leica DM550 fluorescence microscope using Leica Software. Dye separation analysis was done using Leica Software. Two-photon imaging was performed following the reported protocols¹⁴.

Immunohistochemistry. Tissues were fixed in 10% formalin, paraffin-embedded and cut into 4–6 μ m sections. Following dehydration of the slides, antigen retrieval was done in 10 mM citrate buffer pH 6.0 (Na₃H₆O₇). Blocking was done using 0.1% BSA in PBS with 10% normal goat serum for 30 min. Primary antibodies were left overnight at 4°C. The following primary antibodies were used to perform staining: E-cadherin (BD Biosciences, 610181); β -catenin (BD biosciences, 610153); Her2 (Abcam, ab2428). VectaStain Elite ABC Rabbit IgG (PK-6101) and Mouse IgG (PK-6102) kits from Vector Laboratories were used for secondary antibodies. Secondary antibodies were left for 1 h at room temperature. The DAB substrate kit (Vector Laboratories, SK-4100) was used for enzymatic substrate. Mounting was done using Vectashield mounting media (Vector Laboratories, H-1400). Tissues were fixed in 10% formalin, paraffin-embedded and cut in 4–6 μ m sections. For BALB-Her2-T staining, paraffin embedded BALB/c and BALB-Her2-T mammary gland and tumours sections were a gift from C. Klein. Following dehydration of the slides, antigen retrieval was done in 10 mM citrate buffer (Na₃H₆O₇). Triton X-100 was used to permeabilize cells and blocking was done using 0.1% BSA in PBS with 10% normal goat serum or normal donkey serum (Sigma-Aldrich D9663). Primary antibodies were left overnight at 4°C and secondary antibodies were left for 1 h at room temperature. The following primary antibodies were used: E-cadherin (BD Biosciences, 610182); β -catenin (BD biosciences, 610153); Her2 (Abcam, ab2428); p-Atf2 (Cell Signaling, 9226); p-Rb ((Ser249/Thr252), Santa Cruz, sc-16671); CK8/18 (Progen 412121); and p-p38 (BD Biosciences, 612281). The following secondary antibodies were used: AlexaFluor 488 goat-anti-mouse, AlexaFluor 568 goat-anti-rabbit, AlexaFluor 488 donkey-anti-goat, AlexaFluor 547 donkey-anti-rabbit. Slides were mounted using Prolong Antifade mounting media with DAPI (Molecular Probes P36930).

Western blot. Samples were collected in $1\times$ RIPA buffer and centrifuged at 4°C, 15,000g to clarify the lysate. Protein concentrations were calculated using the BioRad Protein Assay Dye Reagent (BioRad 500-0006) and a standard BSA curve. Samples were then boiled for 5 min at 95°C in sample buffer (0.04 M Tris-HCl pH 6.8, 1% SDS, 1% β -mercaptoethanol and 10% glycerol). 6–10% SDS-PAGE gradient gels were run in running buffer (25 mM Tris, 190 mM glycine, 0.1% SDS) and transferred to PVDF membranes in transfer buffer (25 mM Tris, 190 mM glycine, 20% methanol). Membranes were then blocked in 5% milk in

TBST (Tris-buffered saline containing Tween-20) buffer. Primary antibodies were left overnight at 4 °C. Following washing with TBST buffer HRP-conjugated secondary antibodies were left at room temperature for 1 h. Western blot development was done using Amersham ECL Western Blot Detection (GE, RPN 2106) and GE ImageQuant LAS 4010. Primary antibodies used were: p-p38 (BD Biosciences, 612281); p38 (BD Biosciences, 612169); p-ERK1/2 (Cell Signaling, 9101); ERK (BD Biosciences, 610031); and Gapdh (Calbiochem, CB1001). Secondary antibodies used were peroxidase horse-anti-mouse IgG (Vector Laboratories, PI2000); biotinylated goat-anti-rabbit IgG (Vector Laboratories, BA-1000).

EMT qPCR array. RNA was extracted from 3D cultures using TriZol extraction as per the manufacturer's recommendations. cDNA was synthesized using the Qiagen RT2 First Strand Kit (Qiagen 330401). Expression of EMT genes was measured using the Qiagen human Epithelial to Mesenchymal Transition PCR array (Qiagen PAHS-090Z) and RT2 qPCR Master Mixes (Qiagen 330521). Plates were run in an ABI PRISM 7900HT sequence detection system. Results were analysed using web-based PCR-array data-analysis software provided by Qiagen. Heat maps were generated based on triplicate runs of the array using GENE-E software (Brode Institute).

Quantitative PCR. RNA was extracted from 2D- and 3D-cell cultures using TriZol following the provider's recommendation. For pre-malignant mammary gland tissue, RNA was extracted using Qiagen's RNeasy Lipid Tissue Midi Kit (Qiagen 74804). 2 µg RNA was reverse transcribed into cDNA using MMuLV Reverse Transcriptase (New England Biolabs, M0253L), MMuLV buffer (NEB, B0253S) and RNase inhibitor (Ambion, AM2682). Quantitative real time-PCR (qRT-PCR) was performed using a Sybr Green Powder PCR Master Mix (Bio-Rad, 170-8882) or by preparing a master mix using Sybr Green (Sigma-Aldrich, S9430), MgCl₂ (NEB, B9021S), dNTPs (NEB N0447L) and Taq DNA polymerase (Sigma-Aldrich D6558) in a Biorad thermocycler. *Gapdh* was used as a housekeeping control for all plates. **qPCR primers.** Human primers; *GAPDH* forward: 5'-GGTGAAGGTCGGAGTCAACGG-3', *GAPDH* reverse: 5'-ATGAAGGGTCATTTGATGGCAACAA-3'; *CDH1* forward: 5'-ATGGGGTCTTGCTATGTTGC-3', *CDH1* reverse: 5'-AAGGCAGAAGGATTGCTTGA-3'; *Twist1* forward: 5'-GTCCGCAGTCTTACGAGGAG-3', *Twist1* reverse: 5'-CCAGCTTGAGGGTCTGAATC-3'; *SNAI1* forward: 5'-AGAGCTGACCTCCCTGTCA-3', *SNAI1* reverse: 5'-TGAAGTAGAGGAGAAGGACGAA-3'; *WNT11* forward: 5'-CATGGAGCTCTGCTTGTGAA-3', *WNT11* reverse: 5'-GCTTCCAAGTGAA GGCAAAG-3'; *WNT5A* forward: 5'-GAAATGCGTGTGGGTGA-3', *WNT5A* reverse: 5'-AGGCATGGGTTTCCATTCT-3'; *WNT5B* forward: 5'-CCAAAGGA TCAGAGGACAG-3', *WNT5B* reverse: 5'-CTCGTTGTTTGCAGGTTCA-3'; *ERBB3* forward: 5'-GGCGGCACTTTTCTCTACTG-3', *ERBB3* reverse: 5'-CGTTCCAAGTATCGCCTCAT-3'; *FZD7* forward: 5'-TGGGTTAATTTTCAFFTC-3', *FZD7* reverse: 5'-GCAGTACGGGAGGAAAAACA-3'; *AXIN2* forward: 5'-CTGGTGCAAAGACATAGCCA-3', *AXIN2* reverse: 5'-GTCCAGCAAACTCTGAGGG-3'. Mouse primers; *Gapdh* forward: 5'-AACTTTGGCATTTGTGGAAGGGCTC-3', *Gapdh* reverse: 5'-TGGAAGA GTGGGAGTTGCTGTTGA-3'; *Twist1* forward: 5'-AACTGGCCT GCAAATCATA-3', *Twist1* reverse: 5'-ACACCGGATCTATTTCATT-3'.

Two-photon intra-vital microscopy of mammary glands. For this work a single laser source tuned to 880 nm provided excitation for both the CFP tumour cells and the 155 kDa TRITC-dextran vascular label, in addition to producing a second harmonic generation signal from collagen fibres. Imaging was done with a 25× 1.05 NA (XLPL25XWMP2, Olympus) water-immersion objective lens so as to bridge between low-magnification visualization of the ductal tree and high resolution single-cell imaging. For each mouse, a large 25–100 field mosaic was acquired to ascertain the ductal tree structure, from which three separate fields were selected for time lapse imaging, using 5 µm z steps to a depth of approximately 50 µm, with each stack taken approximately every 2 min, for 4–6 h. Images were reconstructed and analysed either in ImageJ³⁴, using the custom written ImageJ plugin, ROI_Tracker¹⁴ or with Imaris (Bitplane). Mice were anaesthetized using 0.75–2.5% isoflurane, depilated and a skin flap surgery performed exposing the 2nd and 3rd mammary fat pad. The absence of a solid tumour necessitated the development of a custom fixturing technique wherein the exposed fat pad was affixed with cyanoacrylate glue to the edge of a 15-mm window fitted with a 12-mm diameter cover glass. The window was captured on a fixturing plate placed on the microscope xy stage and imaging was performed in the centre of the window away from the glue. Animals were maintained at physiological temperatures throughout imaging with an AirTherm ATX forced-air heater (WPI Inc.) and supplemented intravenously with 50–100 µl of PBS per hour. Intravital imaging was performed using a custom-built two-laser multiphoton microscope following previously reported imaging protocols¹⁷. All procedures were conducted

in accordance with the National Institutes of Health regulations and approved by the Albert Einstein College of Medicine animal use committee. For computational rendering of the videos in Supplementary Video 5 and 8, the signals within the segmented vessel and the manually outlined cell were separately extracted into a sequence of tiff images and then imported into Imaris. A colocalization algorithm was performed on the two signals to identify overlapping pixels. Intensity based surface reconstructions of the vessel (red), tumour cells (cyan) and the colocalization signal (yellow) were created and then animated. Any residual xy drift not eliminated by the fixturing window was removed with post-processing using the StackReg plugin³⁵ for ImageJ.

Experimental metastasis assays. Mammary epithelial cells derived from early lesions (16-week-old MMTV–Her2 females) or tumour-derived cells were intravenously injected (10,000 cells per animal) into nude mice. One month later, mice were euthanized and the number of metastatic foci were counted by histology.

Sphere-forming assays. Mammary epithelial cells from early lesions or from tumours were isolated and seeded in low attachment 24-well plates in 1 ml sphere medium plus 1% methylcellulose (1,000 cells per well, $n = 2$ animals per group, sextuple). The number of spheres (clusters of more than 5 cells) per group was counted one week later. These spheres were then dissociated with trypsin and re-plated and the number of spheres per group was counted one week later.

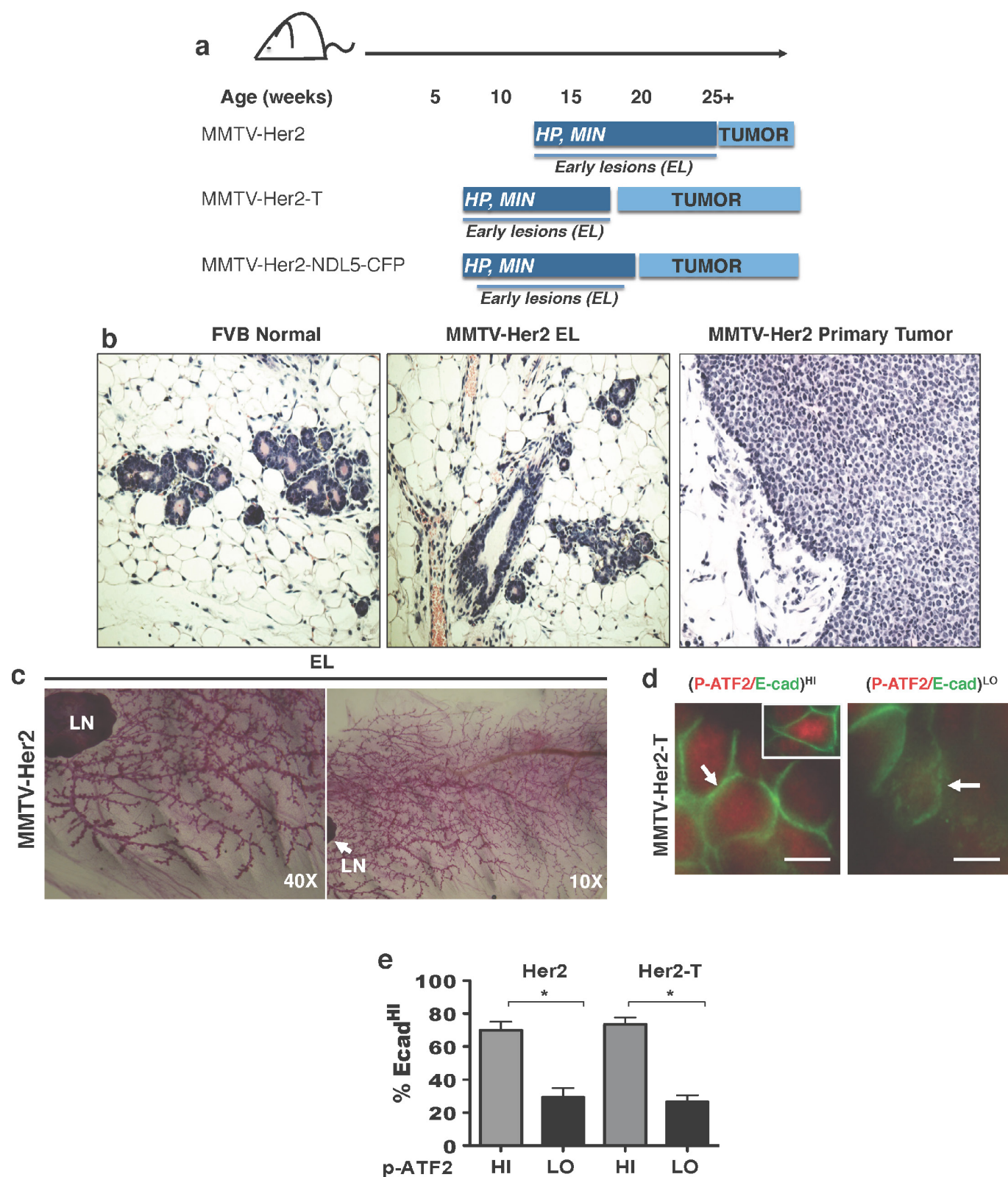
Patient samples. Paraffin-embedded sections from tumours of patients with DCIS or invasive breast cancer were obtained from the Cancer Biorepository at Icahn School of Medicine at Mount Sinai, New York, New York. Samples were de-identified and obtained with Institutional Review Board approval, which indicated that this work does not meet the definition of human subject research according to the 45 CFR 46 and the Office of Human Subject Research. Immunofluorescence and immunohistochemistry analysis was done using samples from 10 DCIS and 20 invasive breast-cancer patients. Invasive breast cancer samples included luminal A, B and HER2-positive subtypes.

Circulating Cancer Cells and DCCs detection. 16-week-old MMTV–Her2 mice were treated with SB203580 (10 mg kg⁻¹) or DMSO for 2 weeks and blood (approximately 500 µl per mouse) was drawn by cardiac puncture following IACUC protocols. Circulating cancer cells (CCCs) were purified using a negative lineage cell-depletion kit (130-090-858, Miltenyi Biotec), fixed and stained with anti-CK8/18 antibody in cytospin preparations. CCCs were counted per ml of blood. Bone-marrow cells from 4 long bones (2 tibiae and 2 femurs per mouse) were flushed out with Minimum Essential Medium Eagle (MEME) (Sigma-Aldrich) using a 26-gauge needle and 1 ml syringe. Tumour cells were enriched by Ficoll-Paque plus (GE Healthcare) density gradient separation and filtered through a 70-µm nylon mesh to remove large aggregates. Cells were fixed with 3% PFA for 20 min on ice and cytospin preparations were carried out by centrifugation of bone-marrow cells at 500 r.p.m. for 3 min using poly-L-lysine-coated slides (Sigma-Aldrich). Bone-marrow-derived DCCs were stained with anti-CK8/18 and anti-Her2 antibodies and cytospin preparations were analysed. We screened $0.5\text{--}2.0 \times 10^6$ bone-marrow cells, which represents 20% of the total bone-marrow cells obtained from 2 tibiae and 2 femurs per mouse after Ficoll gradient separation and then normalized to the total volume in which each bone-marrow DCC sample was resuspended (approximately 1 ml). Similar to mammosphere preparations, all 5 pairs of glands in mice were checked for the presence of any visible or palpable tumours when processed for early cancer cells and none were found as described above.

Statistical analysis. Statistical analysis was done using Prism Software. Differences were considered significant if $P < 0.05$. For most cell culture experiments, one-tailed Student's *t*-tests were performed unless specified. For mouse experiments one-tailed Mann–Whitney *U*-tests were used. Sample sizes were chosen empirically and no exclusion criteria were applied. The investigators were not blinded to allocation during experiments and outcome assessment.

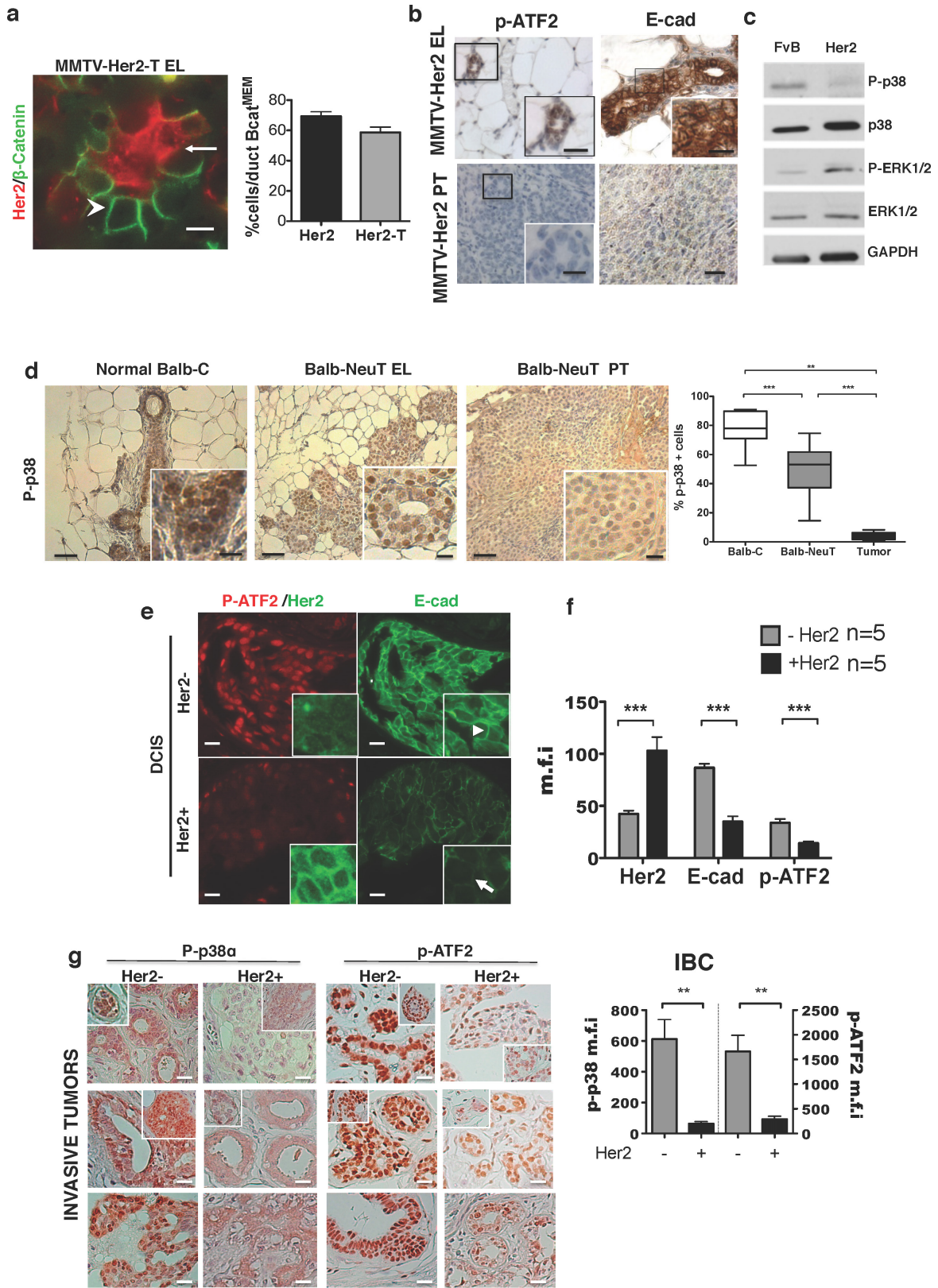
Data availability. The datasets generated during and/or analysed during the current study are available within the paper (and its Supplementary Information) and/or from the corresponding author on reasonable request.

32. Debnath, J., Muthuswamy, S. K. & Brugge, J. S. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* **3**, 256–268 (2013).
33. Debnath, J. & Brugge, J. S. Modelling glandular epithelial cancers in three-dimensional cultures. *Nat. Rev. Cancer* **5**, 675–688 (2015).
34. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
35. Thévenaz, P., Ruttimann, U. E. & Unser M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* **7**, 27–41 (1998).



Extended Data Figure 1 | Progression and staging of MMTV-Her2 mouse models. **a**, Cartoons depicting the three MMTV-Her2 models used in this study and the different time frames for early lesions (EL) and overt primary tumour (PT) development. HP, hyperplasia; MIN, mammary intraepithelial neoplasia. **b**, Haematoxylin and eosin staining for sections of normal FVB mouse mammary tissue, and FVB MMTV-Her2 early lesions or primary tumours. **c**, Whole mounts from mammary glands of FVB MMTV-Her2 mice at the time early lesions were studied. LN, lymph

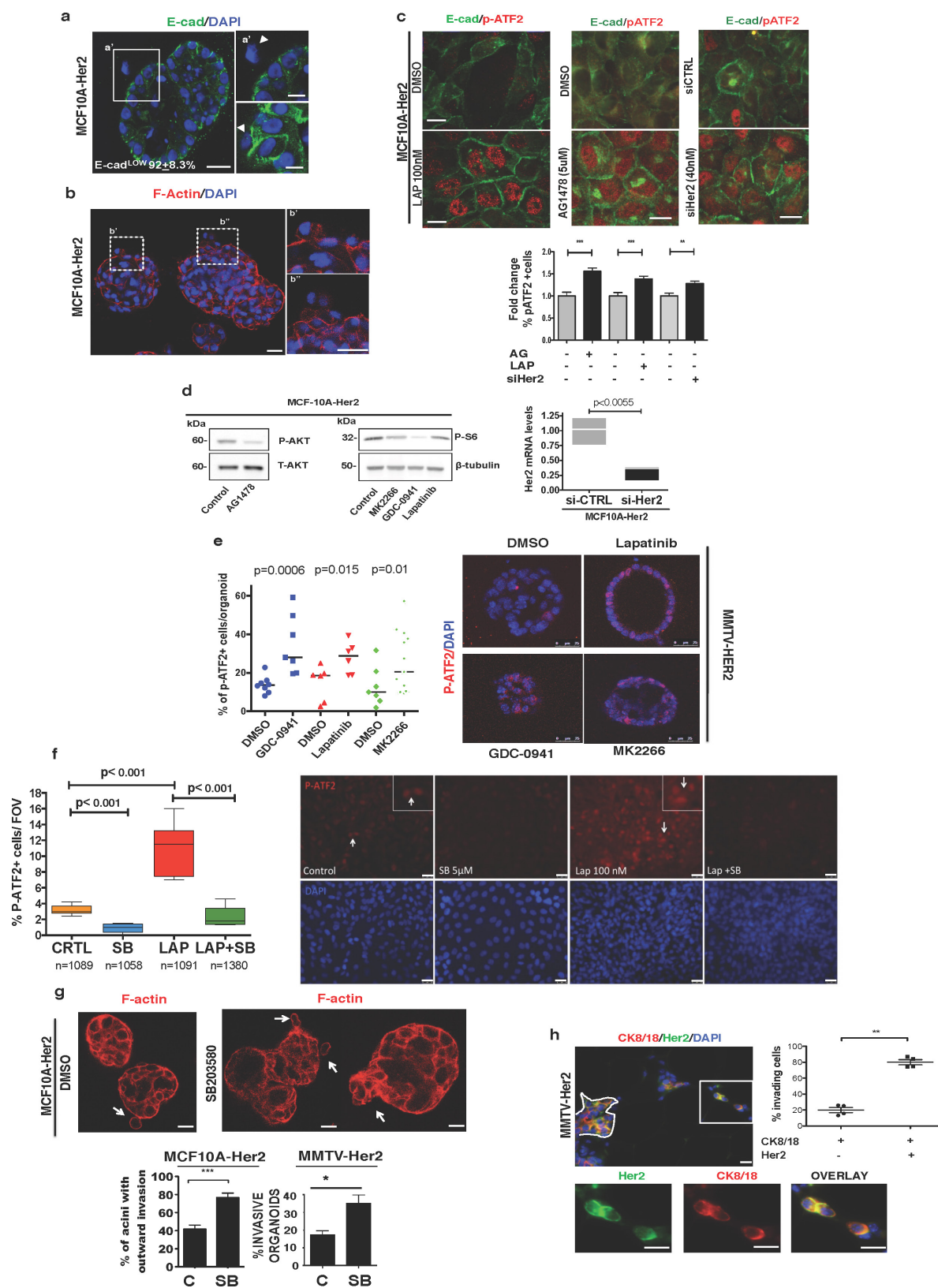
node. **d**, Representative images of E-cad^{hi}p-ATF2^{hi} (left and inset) and E-cad^{lo}p-ATF2^{lo} (right) ducts in the MMTV-Her2-T model. Scale bar, 10 μ m. Arrow in left image, intact E-cadherin junction; arrow in right image, dismantled E-cadherin junction. **e**, Quantification of the percentage of E-cad^{hi} cells per duct that showed high or low p-ATF2 expression in MMTV-Her2 and MMTV-Her2-T models. * $P < 0.01$; one-sided, unpaired t -test; mean \pm s.e.m. (Her2, $n = 30$ ducts; Her2-T, $n = 10$ ducts).



Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | p38 signalling in MMTV-Her2 models and DCIS patient samples. **a**, Early stage MMTV-Her2-T (BALB-NeuT 15 weeks of age) early lesion sections stained for Her2 and β -catenin. Arrows, Her2⁺ β -cat^{MEM-lo} early lesion cells; arrowheads, Her2⁻ β -cat^{MEM-hi} cells. Scale bars, 10 μ m. The digital dye separation module (Leica) was used on the images. Graph, quantification of the percentage of cells per duct with β -cat^{MEM} for both MMTV-Her2 and MMTV-Her2-T models. ($n = 7$ ducts). **b**, Immunohistochemistry for p-ATF2 and E-cadherin in MMTV-Her2 early lesion tissues (age, 14–18 weeks) and primary tumour sections. Boxed regions are magnified in the bottom right panel. Note the loss of both p-ATF2 and E-cadherin in primary tumour samples. Scale bar, 25 μ m. **c**, Western blot for the indicated antigens in lysates of mammary epithelial cells isolated from normal mammary glands (FVB) and tumour cells isolated from MMTV-Her2 overt primary tumours (Her2). GAPDH was used as a loading control. For gel source data, see Supplementary Fig. 1. **d**, Immunohistochemistry for p-p38 in normal epithelium (BALB/c), early lesion tissues (BALB-NeuT early lesions, 7 weeks) and overt primary tumours (BALB-NeuT primary tumours). Graph, percentage of p-p38 positive cells in each stage. $n = 11$ –15 ducts, 5 tumours. Scale bars, 20 μ m (inset) and 50 μ m; ** $P < 0.01$; *** $P < 0.0001$; one-tailed Mann–Whitney U -test.

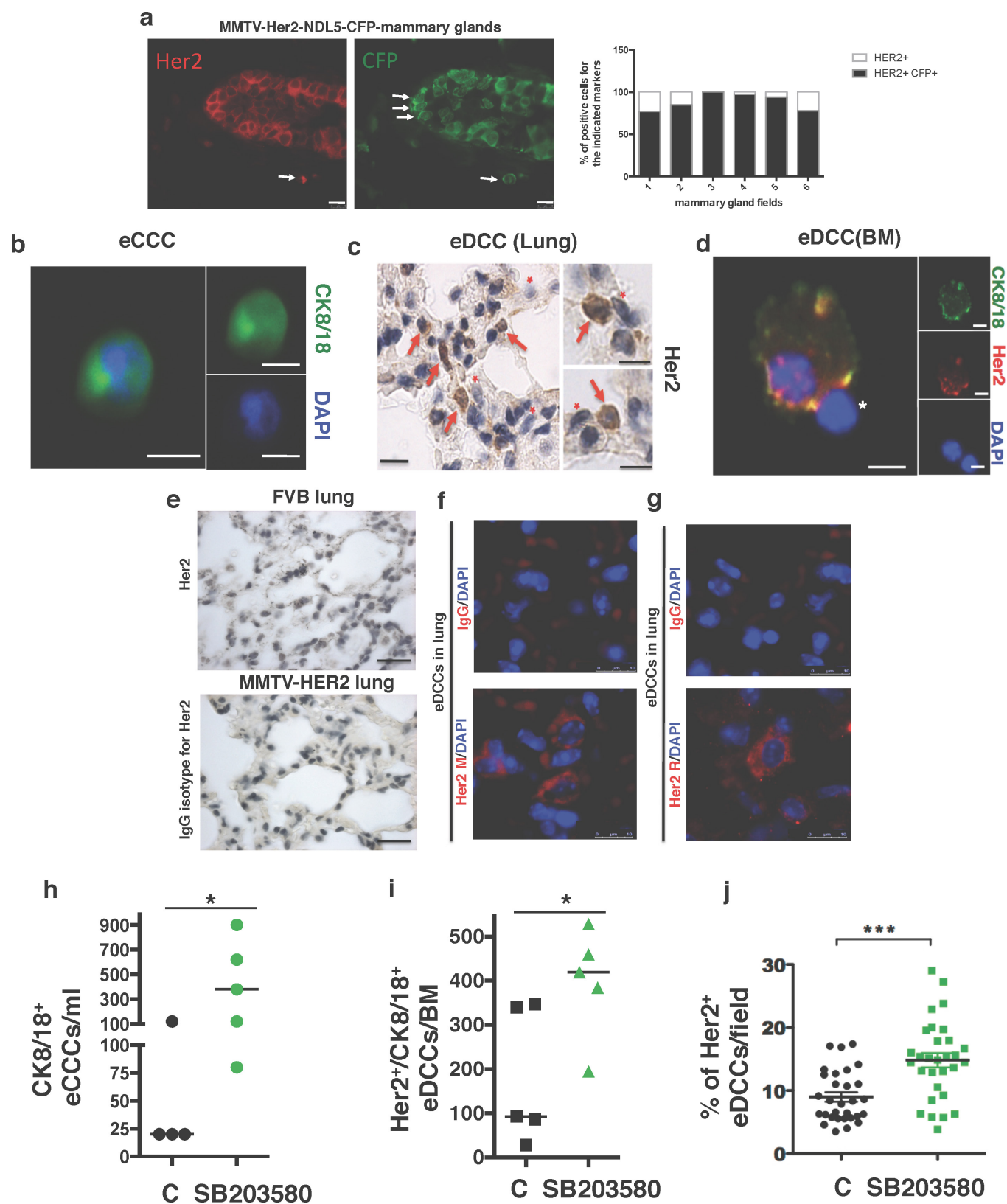
e, Representative images of parallel sections from DCIS patient samples stained for p-ATF2 (red), Her2 (green in insets lower row), or E-cadherin (green in large panels and insets upper row). Samples were Her2-positive ($n = 5$) or -negative ($n = 5$) by immunofluorescence microscopy analysis for Her2 (insets top and bottom row left, green). Inset right column, detail of E-cadherin junctions in Her2⁺ and Her2⁻ samples. Arrowhead, strong E-cadherin junctions; arrow, weak E-cadherin staining. Scale bars, 25 μ m and 10 μ m (inset). **f**, Metamorph software was used to quantify Her2, E-cadherin and p-ATF2 fluorescence signal intensity in 10 DCIS samples shown in panel **e**. Mean fluorescence intensity (m.f.i.) \pm s.e.m. per cell per field from Her2⁺ (black bars, $n = 5$) compared to Her2⁻ (grey bars, $n = 5$) samples from patients with DCIS. *** $P < 0.05$; two-way ANOVA. **g**, Immunohistochemistry for p-p38 α and p-ATF2 performed on invasive breast cancer (IBC) tumours from patients ($n = 20$). Samples were classified as Her2⁺ ($n = 10$) or Her2⁻ ($n = 10$) by the pathology service. Note the significant reduction in both p-p38 α and p-ATF2 in Her2⁺ tumours. Insets show additional patient samples for each group. Graph, metamorph was used to determine the mean signal intensity \pm s.e.m. per field for p-p38 and p-ATF2. p-p38 intensity, left axis and first two columns of the graph. p-ATF2 intensity, right axis and last two columns of the graph. ** $P < 0.05$; unpaired t -test. Scale bars, 25 μ m.



Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Characterization of invasive and signalling properties of Her2⁺ early lesions. **a**, MCF10A-HER2 organoids stained for E-cadherin (green) and DAPI (blue). Left, a representative organoid. Scale bar, 25 μ m. Right, details of invading E-cad^{lo} cells (top and bottom). Arrowheads, outward invading E-cad^{lo} cells. Scale bar, 10 μ m. Approximately 35–40% of MCF10A-HER2 organoids show outward invasion of one cell per organoid in equatorial sections; $92 \pm 8.3\%$ of those invading cells are E-cad^{lo}. **b**, MCF10A-HER2 organoids were stained for F-actin (red) and DAPI (blue). Note the extensions of F-actin from invading cells (boxed areas and right top and bottom insets) still in contact with the organoid. Scale bars, 20 μ m. **c**, Detection of E-cadherin and p-ATF2 in MCF10A-HER2 cells treated with or without lapatinib (100 nM), AG1478 (5 μ M) and siRNA targeting HER2 (40 nM) for 24 h; E-cadherin (green), pATF2 (red). Graph, fold change of the percentage of p-ATF2⁺ cells. Data are mean \pm s.e.m.; $**P < 0.01$; $***P < 0.001$; one-sided, unpaired *t*-test; $n = 3$ experimental replicates, 10 images per treatment. **d**, MCF10A-Her2 cells were treated for 24 h with AG1478 (1 μ M) left, or with the AKT inhibitor MK2266 (5 μ M), pan-PI3K inhibitor GDC-0941 (1 μ M) or lapatinib (1 μ M), right. Western blots for p-AKT and total AKT (T-AKT) (left) or p-S6 and β -tubulin (right). Gel source data in Supplementary Fig. 1. Graph, control for Her2 knockdown in MCF10A-HER2 cells; one-sided, unpaired *t*-test; median and range are shown. **e**, MMTV-Her2 early lesion organoids were treated with GDC-0941 (1 μ M), lapatinib (1 μ M) or MK2266 (5 μ M) for 24 h. Organoids were fixed and stained for p-ATF2.

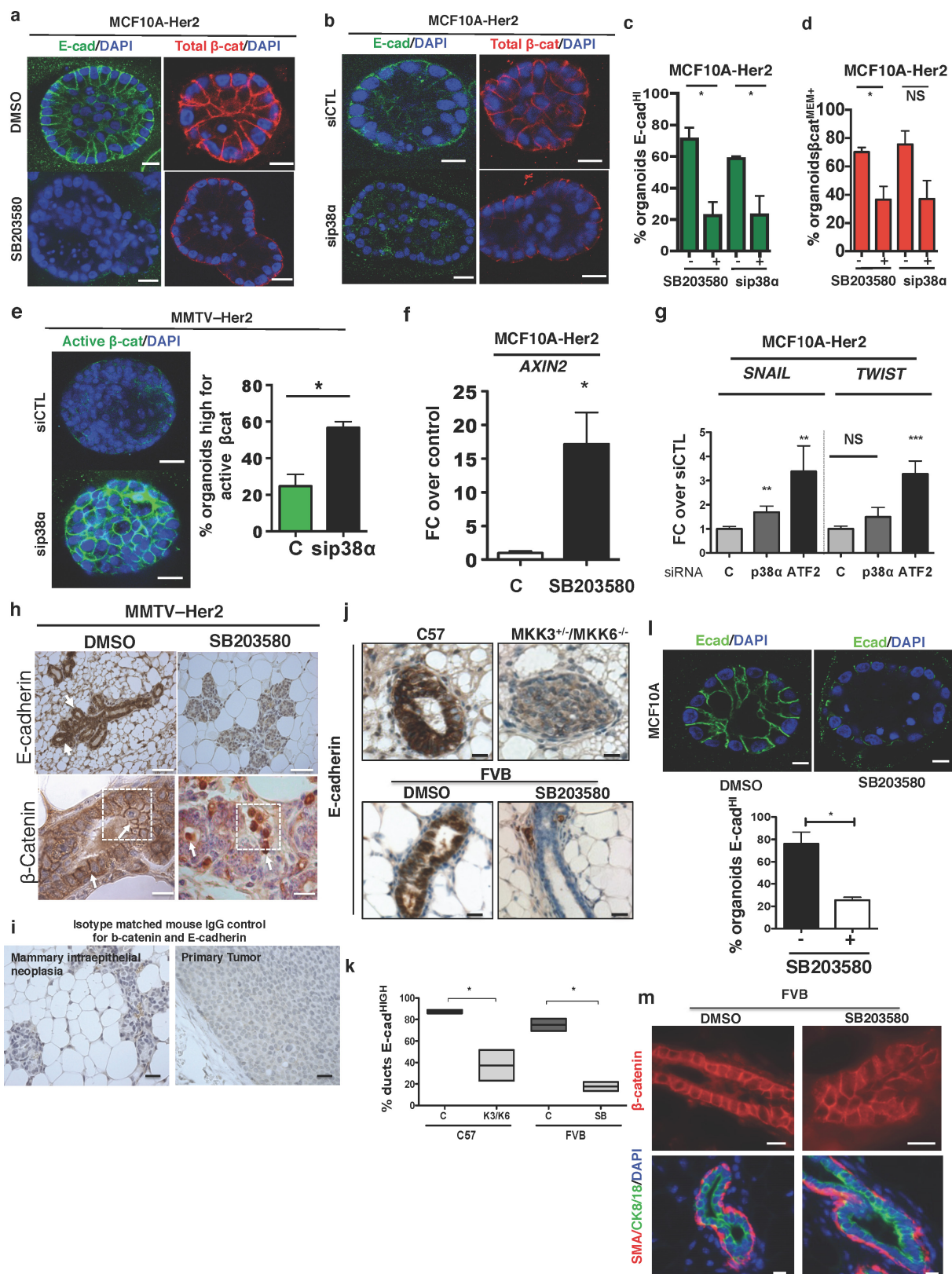
Graph, percentage of p-ATF2⁺ cells per organoid. Scale bars, 25 μ m. Median \pm s.e.m.; one-sided, unpaired *t*-test. **f**, Left, quantification of the percentage of nuclear p-ATF2⁺ MCF10A-HER2 cells treated for 24 h with vehicle (CTRL), 5 μ M SB203580 (SB), 100 nM lapatinib (LAP) or the combination of the two drugs. Right, representative immunofluorescence images of the p-ATF2 signal (red); DAPI (blue) was used to count total cell numbers. Insets and arrows show a detail of nuclear p-ATF2 levels in the respective groups. One-sided Mann-Whitney *U*-test at 95% confidence; median and range are shown, $n = 2$ independent wells per condition; $n > 150$ cells scored per condition. Scale bars, 25 μ m. **g**, MCF10A-HER2 organoids treated for 6 days with SB203580 or DMSO and stained for F-actin (red). Bottom left graph, percentage \pm s.e.m. of MCF10A-HER2 organoids with outward invasion (DMSO $n = 109$; SB203580 $n = 87$) $***P < 0.001$; one-sided, unpaired *t*-test. Scale bars, 10 μ m. Bottom right graph, percentage of invasive MMTV-Her2 organoids (DMSO $n = 11$; SB203580 $n = 9$). $P = 0.01$; one-sided, unpaired *t*-test; mean \pm s.d. Representative of 3 biological replicates. **h**, MMTV-Her2 early lesion sections (age, 14–18 weeks) stained for CK8/18 (red), Her2 (green) and nuclei (DAPI, blue). Top image, a duct is outlined. The boxed region and bottom images show CK8/18⁺ and Her2⁺ singlets or doublets within the stroma near ducts. Graph, percentage of stroma-invading cells that were either double positive for both CK8/18 and Her2 or single CK8/18⁺. Mean \pm s.e.m.; $**P < 0.01$; one-sided Mann-Whitney *U*-test; $n = 4$ mice, 60–80 cells per mouse. Scale bars, 10 μ m.



Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | Quantification of early dissemination by Her2⁺ early cancer cells. **a**, MMTV–Her2–NDL5–CFP early lesion mammary gland tissues (seven-week-old females) co-stained for CFP (green) and Her2 (red). Arrows, co-distribution of Her2 and CFP. Graph, percentage of positive cells for the single Her2 staining (white bars) or double co-staining for Her2 and CFP (black bars) per field. Approximately 88% of early lesion cells are positive for Her2 and CFP. Scale bars, 10 μ m. **b**, Early circulating cancer cells (eCCCs) were detected in cytopsin preparations by staining for CK8/18 (green) and nuclei with DAPI (blue) after a Ficoll gradient and negative selection (see Methods). Scale bar, 10 μ m. **c**, Detection of eDCCs in lung sections from MMTV–Her2 mice by immunohistochemistry for Her2 (rabbit anti-Her2 antibody (Abcam, ab2428)). Scale bar, 25 μ m. Right, augmented images from additional sections. Red arrowheads, Her2-positive DTCs; red asterix, host Her2-negative cells. Scale bars, 10 μ m. Staining controls are shown in **e**. **d**, eDCCs in the bone marrow of MMTV–Her2 mice detected in cytopsin preparations of whole bone-marrow samples after a Ficoll gradient and staining for CK8/18 (green), Her2 (red) and DAPI (blue). CK8/18⁺, Her2⁺ or double-positive cells were considered eDCCs. Right, individual channel

signals. Left, merged channels on the right detecting a bone-marrow CK8/18⁺Her2⁺ DCC (arrow) next to a CK8/18[−]Her2[−] bone-marrow cell (asterix). Scale bars, 10 μ m. **e**, Top, Immunohistochemistry for Her2 in non-transgenic FVB lung sections. Her2⁺ cells were undetectable in FVB lung sections. Bottom, IgG isotype for the Her2 antibody used in **c** in lungs of MMTV–Her2 mice. Scale bars, 50 μ m. **f**, **g**, Top, IgG control images for eDCC detection in MMTV–Her2 lung sections. Bottom, example of Her2⁺ (red) staining using the Calbiochem (OP15L) (**f**) and Abcam (ab2428) (**g**) anti-Her2 antibodies. M, mouse; R, rabbit. Scale bars, 10 μ m. **h**, eCCCs detected by CK8/18⁺ as in **b** in blood of MMTV–Her2 mice (age, 14–18 w) treated for 2 weeks with DMSO (C) or the p38 α / β inhibitor SB203580 (DMSO n = 4; SB n = 5 mice). **i**, eDCCs detected by CK8/18⁺ as in **d** in bone marrow of MMTV–Her2 mice treated as in **h** (n = 5 mice per group). **j**, eDCCs detected in the lung of MMTV–Her2 mice carrying only early lesions as in **c** and treated as in **h**. Graph, percentage of Her2⁺ eDCCs per field in each group (n = 30 fields, 3 mice per treatment). For **h–j**, median and individual fields (**j**) or mice (**h**, **i**); * P < 0.05; *** P < 0.001; one-sided Mann–Whitney U -test.

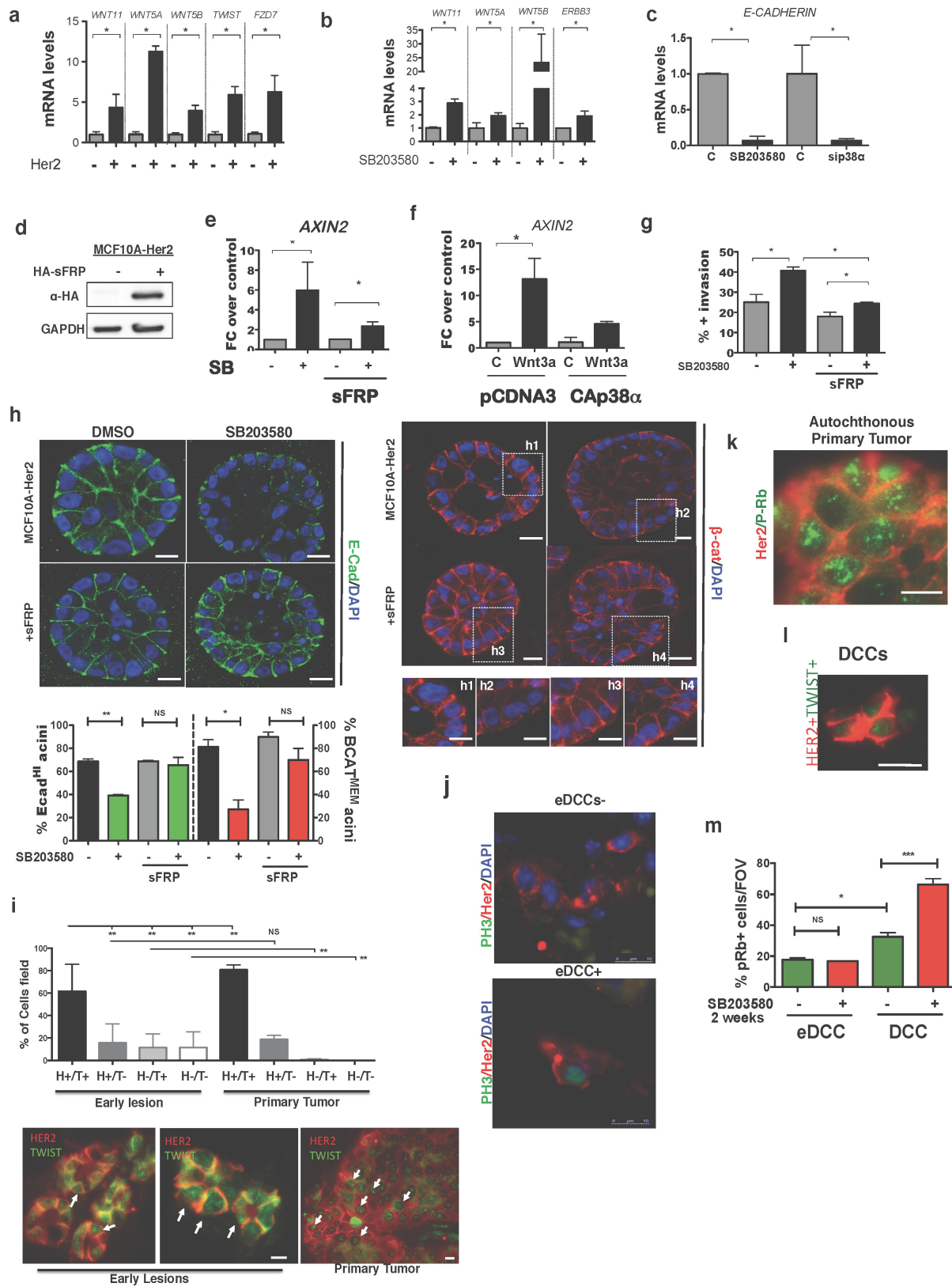


Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | EMT markers in Her2⁺ early lesion cells.

a, MCF10A-HER2 organoids treated for 6 days with SB203580 or control (DMSO) were stained for E-cadherin (green) and β -catenin (red) or DAPI (blue); MCF10A-HER2, $n = 20$ organoids per treatment. Scale bars, 10 μ m (top), 20 μ m (bottom). **b**, MCF10A-HER2 organoids treated with siRNAs targeting *p38a* or a non-targeting control (siCTL); $n = 20$ organoids per treatment. Scale bars, 10 μ m (top), 20 μ m (bottom). **c**, Quantification of **a** and **b**. * $P < 0.048$; NS, not significant; one-sided, unpaired t -test. **d**, Graph, percentage \pm s.e.m. β -catenin^{MEM} in MCF10A-HER2 with/without SB203580 and with/without *p38a* siRNA. * $P = 0.0047$; one-sided, unpaired t -test. **e**, MMTV-HER2 organoids treated with *p38a* or control siRNA (48 h) and stained to detect active β -catenin (see Methods). Graph, percentage \pm s.e.m. of organoids stained for active β -catenin ($n = 10$ organoids per treatment). Scale bars, 25 μ m. $P = 0.0059$; one-sided, unpaired t -test. **f**, *AXIN2* mRNA expression in MCF10A-HER2 cultures treated for 24 h with DMSO control (C) or SB203580 (5 μ M). Technical triplicate determinations were normalized to *GAPDH* and fold change (FC) over control was determined for five biological replicates. $P < 0.05$; one-sided, unpaired t -test, mean \pm s.e.m. **g**, mRNA levels for *SNAIL1* and *TWIST1* normalized to *GAPDH* in MCF10A-HER2 3D cultures treated with siRNA targeting the *p38a* isoform or ATF2 from day 6–12. Graph,

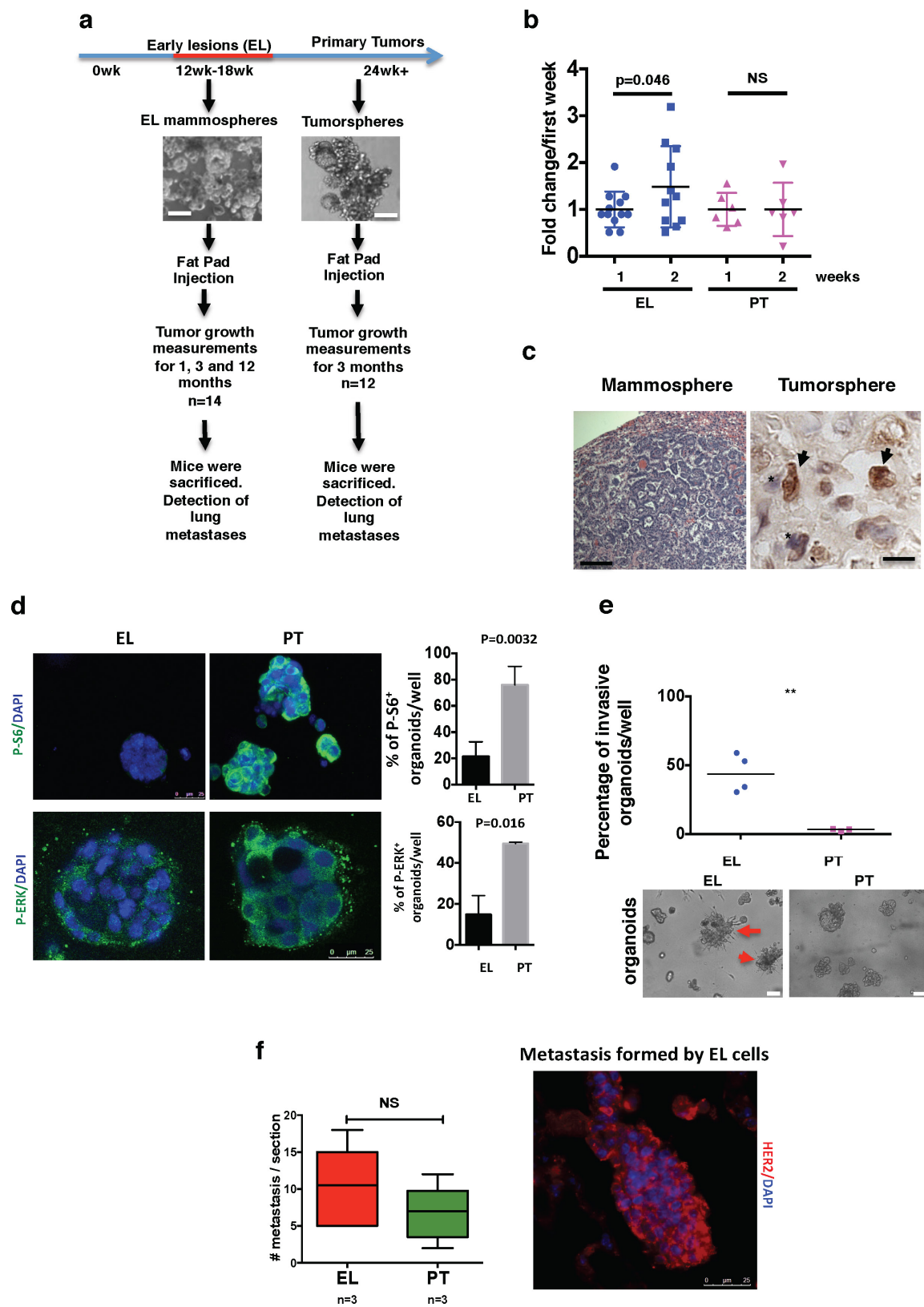
fold change over control in three biological replicates. ** $P < 0.01$; *** $P < 0.0001$; one-sided, unpaired t -test; mean \pm s.e.m. **h**, Sections of MMTV-HER2 early lesions in mice treated for 2 weeks with SB203580 (see Methods) stained for E-cadherin (top) and β -catenin (bottom). Scale bars, 15 μ m. Arrows, membrane E-cadherin, or β -catenin (bottom left) or nuclear β -catenin (bottom right). Boxed area is shown in Fig. 3c. **i**, Isotype-matched mouse IgG control immunohistochemistry for β -catenin and E-cadherin in mammary intraepithelial neoplasia and a primary tumour. Scale bars, 25 μ m. **j**, E-cadherin immunohistochemistry in C57BL/6 (WT) and *Mkk3*^{-/-} *Mkk6*^{+/-} mice¹⁰ or FVB mice treated with SB203580 (see Methods). **k**, Quantification of **j**. Mean \pm s.e.m.; * $P < 0.01$; one-sided, unpaired t -test. **l**, MCF10A organoids were treated for 6 days with control (DMSO) or SB203580 (5 μ M), and fixed and stained for E-cadherin (green). Graph, percentage of E-cad^{hi} organoids in two experiments; 15 organoids per treatment per trial. Scale bars, 10 μ m. Mean \pm s.e.m.; * $P < 0.01$; one-sided, unpaired t -test. **m**, Top, immunofluorescence for β -catenin (red) on mammary gland sections of DMSO- or SB203580-treated FVB mice (see Methods). Scale bars, 25 μ m. Bottom, immunofluorescence for α -smooth muscle actin (SMA, red), CK8/18 (green) and DAPI (blue) on the same tissues. Scale bars, 20 μ m.



Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | Wnt signalling in Her2⁺ organoids and eDCC characterization. **a**, qPCR confirmation of EMT genes identified in Fig. 3g comparing MCF10A and MCF10A-HER2 organoids. Mean \pm s.e.m. shown as fold change over control. Values normalized to *GAPDH* from triplicate samples. * $P < 0.05$; one-sided, unpaired *t*-test. **b**, qPCR confirmation of genes identified in Fig. 3g in MCF10A-HER2 organoids treated with DMSO or SB203580. Mean \pm s.e.m. shown as fold change over control. Values normalized to *GAPDH* from triplicate samples. * $P < 0.05$; one-sided, unpaired *t*-test. **c**, qPCR for *CDH1* mRNA in MCF10A-HER2 organoids treated for 6 days with SB203580 (5 μ M) or *p38a* siRNA (20 nM). Fold change over control for biological triplicates. DMSO, control for SB203580 and scrambled siRNA, control for *p38a* siRNA. Mean \pm s.e.m.; * $P < 0.05$; one-sided, unpaired *t*-test. **d**, Western blot for haemagglutinin (HA)-tagged SFRP1 constructs in MCF10A-HER2-SFRP1 cell lines. Gel source data, see Supplementary Fig. 1. **e**, *Axin2* mRNA levels in MCF10A-HER2 and MCF10A-HER2-SFRP1 cells treated with or without SB203580 (5 μ M) for 24 h. Fold change over control; error bars denote s.e.m. for biological sextuplicates. * $P < 0.05$; one-sided, unpaired *t*-test. **f**, *Axin2* mRNA levels measured in MCF10A cultures transfected with pcDNA3 (empty vector) or CA-p38 α (D176A and F372S mutant) plasmids and then treated with or without WNT3A for 24 h. Fold change over control is shown; error bars denote s.e.m. for biological triplicates. * $P < 0.02$; one-sided, unpaired *t*-test. **g**, Percentage of outward-invading cells from MCF10A-HER2 and MCF10A-HER2-SFRP1

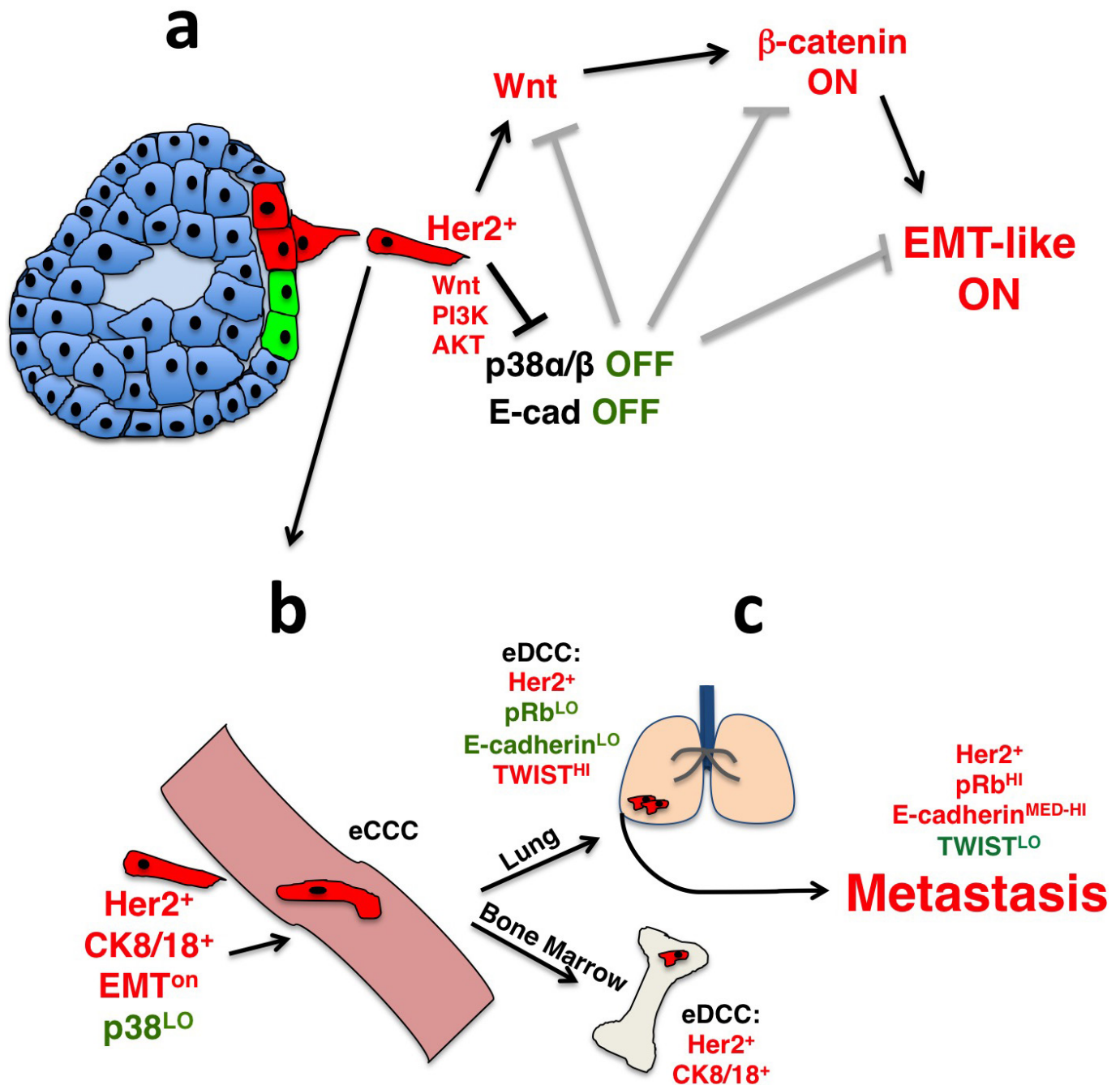
organoids treated for 6 days with DMSO or SB203580 (5 μ M). $n = 20$ organoids per treatment, biological duplicates; data are shown as mean \pm s.e.m.; * $P < 0.05$; one-sided, unpaired *t*-test. **h**, Left, E-cadherin (green) in MCF10A-HER2 and MCF10A-HER2-SFRP1 organoids treated for 6 days with SB203580 (5 μ M). Right, β -catenin (red) in organoids treated as on the left. Insets (h1–h4) show magnified boxed regions. Graph, percentage of E-cad^{hi} (green bars, left axis) and β -catenin^{MEM} (red bars, right axis) organoids. Error bars denote s.e.m.; NS, not significant; * $P < 0.003$; ** $P < 0.02$; one-sided, unpaired *t*-test; $n = 20$ organoids per treatment, biological duplicates. Scale bars, 10 μ m. **i**, Quantification of early-lesion or primary-tumour cancer cells with the indicated profiles; 4 animals per group. ** $P < 0.01$; one-sided Mann–Whitney *U*-test; mean \pm s.e.m. Bottom, immunofluorescence for Twist1^{hi} (T⁺) protein in HER2⁺ (H⁺) cancer cells in early lesions ($n = 883$ cells) or primary tumours ($n \geq 3,000$ cells). **j**, Immunofluorescence for p-H3 (green) and Her2 (red) in eDCCs from MMTV–Her2 lung sections. Scale bars, 10 μ m. **k**, Immunofluorescence for Her2 (red) and p-Rb (green) in spontaneous primary MMTV–Her2 tumours. Scale bar, 10 μ m. **l**, Representative image of Her2⁺Twist1⁺ lung DCCs from 33-week-old MMTV–Neu mice. $n = 500$ cells, 4 animals per group. Quantification shown in Fig. 4f. Scale bar, 10 μ m. **m**, Percentage of Her2⁺ and p-Rb⁺ cells per field of view (FOV) in MMTV–Her2 mice treated as in Fig. 3c. Lungs sections from 3 animals. * $P < 0.02$; *** $P < 0.0001$; one-sided, unpaired *t*-test; error bars represent \pm s.e.m.



Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Metastasis-initiating potential of HER2⁺ early lesion cells. **a**, Experimental approach for testing tumorigenic and metastatic potential. Early lesions cells from mouse mammary glands (age, 12–18 weeks) and primary tumour cells were seeded in mammosphere medium (see representative images). Approximately 300 mammospheres were injected into the fat pad of nude mice. Primary tumour formation and metastasis was monitored for 1, 3 and 12 months (mammospheres group) or for 3 months (tumourspheres group). Primary tumour and metastasis incidence are shown in Fig. 4g. **b**, Sphere-forming efficiency for early lesion cancer cells (age, 16 weeks) and primary tumour cancer cells. After one week (1) in culture, spheres were disaggregated and replated to test self-renewal capacity for another week (2). $n = 6$ replicates; one-sided, unpaired t -test; data are mean \pm s.d. Representative of 3 biological replicates. **c**, Left, haematoxylin and eosin staining of lung macro-metastasis in nude mice injected with MMTV–Her2 early lesion mammospheres. Scale bar, 200 μ m. Right, immunohistochemistry for Her2⁺ DCCs in mice injected with tumourspheres. Scale bar, 10 μ m.

Arrows, Her2⁺ DCCs; asterisks, Her2[−] cells. **d**, Left, immunofluorescence detection of p-ERK1/2 and p-S6 in organoids produced by MMTV–Her2 early lesion or primary tumour cells. Right, percentage of p-S6 or p-ERK⁺ organoids per well. $n =$ triplicates; one-sided, unpaired t -test; mean \pm s.e.m. **e**, Mammosphere from early lesion cells or tumourspheres from primary tumour cells were directly embedded in 3D Matrigel to monitor organoid behaviour for 3 days. Top, percentage of invasive organoids in each group. EL, early lesion; PT, primary tumour. Bottom, representative images used to quantify the invasive nature of early lesion mammospheres (left) compared to primary tumour tumourspheres (right). $P < 0.0021$; one-sided, unpaired t -test; mean \pm s.e.m. **f**, Early lesion and primary tumour single-cell suspensions were injected intravenously (tail vein) in nude mice (50,000 cells per animal). Lungs were collected (after 4 weeks) and processed for haematoxylin and eosin, and immunofluorescence for Her2 detection (right). Graph, number of metastatic nodules per section per animal lung ($n = 3$ mice in each group). NS, not significant; Mann–Whitney U -test; median and range are shown.



Extended Data Figure 9 | Cartoon depicting the mechanism of early dissemination by Her2⁺ early lesion cells. **a**, Early Her2⁺ early lesion cancer cells (red) turn on Wnt, PI3K and AKT signalling, inhibit p38 activation and E-cadherin-junction formation allowing for a Twist1^{hi} EMT-like invasive program; p38 and E-cadherin inhibit the Wnt- and β -catenin-driven EMT-like program and invasion (grey inhibitory symbols). **b**, Her2⁺-p38^{lo}Twist1^{hi}E-cad^{lo} early lesion cancer cells, which

retain CK8/18 expression can intravasate and disseminate. **c**, In lungs more than 85% of eDCCs (red) were Her2⁺E-cad^{lo}(p-Rb or p-H3)^{lo}, suggesting a large population of dormant cells. Most eDCCs are also Twist1^{hi}E-cad^{lo}. Nevertheless, eDCCs can initiate metastasis, which correlated with the acquisition of a Twist1^{lo}E-cad^{med-hi} phenotype. In the bone marrow, eDCCs were Her2⁺CK8/18⁺ and remain dormant for the duration of the experiments, as bone lesions were never observed.

Extended Data Table 1 | Quantification of eCCCs and eDCCs in MMTV-Her2 mice

Disseminating population	Incidence	MMTV-Her2 Median (Min-Max)
eCCC	100%	20 (20-120) CCCs/ml
eDCC BM	100%	86.54 (28-350) / BM (4 hind limb long bones)
eDCC lung	100%	8.1 (3.5-17.4) %DCCs/field

eCCCs, as well as lung and bone-marrow eDCCs, could be detected in 100% of mice (column two) ($n=4$) (for images, see Extended Data Fig. 4h–j). The median number of CK8/18⁺ eCCC per ml of blood from cytopsin preparations of whole blood samples is shown with range ($n=4$). Bone-marrow eDCCs were detected as CK8/18⁺, Her2⁺ and co-stained for both markers in cytopsin preparations from 4 hindlimb long bones (2 tibias, 2 femurs) and the number of eDCCs along with minimum and maximum values in the whole bone marrow is shown ($n=5$). The median percentage of lung eDCCs per field detected by immunohistochemistry of lungs is shown with minimum and maximum number of DCCs ($n=30$ fields, $n=3$ animals). The median number of eDCCs in the bone-marrow compartment is equivalent to approximately 20.2 eDCCs per 10^6 bone-marrow host cells in control groups and approximately 46 eDCCs per 10^6 bone-marrow host cells in SB203580-treated animals.

mRNA quality control is bypassed for immediate export of stress-responsive transcripts

Gesa Zander^{1*}, Alexandra Hackmann^{1*}, Lysann Bender^{1*}, Daniel Becker¹, Thomas Lingner², Gabriela Salinas² & Heike Krebber¹

Cells grow well only in a narrow range of physiological conditions. Surviving extreme conditions requires the instantaneous expression of chaperones that help to overcome stressful situations. To ensure the preferential synthesis of these heat-shock proteins, cells inhibit transcription, pre-mRNA processing and nuclear export of non-heat-shock transcripts, while stress-specific mRNAs are exclusively exported and translated¹. How cells manage the selective retention of regular transcripts and the simultaneous rapid export of heat-shock mRNAs is largely unknown. In *Saccharomyces cerevisiae*, the shuttling RNA adaptor proteins Npl3, Gbp2, Hrb1 and Nab2 are loaded co-transcriptionally onto growing pre-mRNAs. For nuclear export, they recruit the export-receptor heterodimer Mex67–Mtr2 (TAP–p15 in humans)². Here we show that cellular stress induces the dissociation of Mex67 and its adaptor proteins from regular mRNAs to prevent general mRNA export. At the same time, heat-shock mRNAs are rapidly exported in association with Mex67, without the need for adaptors. The immediate co-transcriptional loading of Mex67 onto heat-shock mRNAs involves Hsf1, a heat-shock transcription factor that binds to heat-shock-promoter elements in stress-responsive genes. An important difference between the export modes is that adaptor-protein-bound mRNAs undergo quality control, whereas stress-specific transcripts do not. In fact, regular mRNAs are converted into uncontrolled stress-responsive transcripts if expressed under the control of a heat-shock promoter, suggesting that whether an mRNA undergoes quality control is encrypted therein. Under normal conditions, Mex67 adaptor proteins are recruited for RNA surveillance, with only quality-controlled mRNAs allowed to associate with Mex67 and leave the nucleus. Thus, at the cost of error-free mRNA formation, heat-shock mRNAs are exported and translated without delay, allowing cells to survive extreme situations.

To investigate how regular mRNAs are retained in the nucleus during stress, we analysed whether Mex67 and its adaptor proteins or the loading factors Yra1, and Hpr1 of the THO complex, which helps couple mRNA transcription with processing and export into the cytoplasm², dissociate from mRNAs. RNA co-immunoprecipitation (RIP) experiments revealed that this is indeed the case, while the splicing factor Prp17 remained bound^{3,4} (Fig. 1a and Extended Data Fig. 1a). Gbp2 required shorter time-shifts because it aggregates reversibly⁵ (Extended Data Fig. 1b). The dissociation of Mex67 and its adaptors is not complete, because dissociation occurs mainly in the nucleus. The cytoplasmic mRNA pool is excluded from translation as it accumulates with bound proteins in stress granules⁶. It is likely that Mex67 adaptor proteins dissociate to remove the export receptor from the mRNA, as the interaction between the two did not change upon stress (Fig. 1b and Extended Data Fig. 1c).

To identify the requirements of heat-shock mRNA export, we conducted fluorescence *in situ* hybridization (FISH) experiments and found export defects in *mex67-5* mutants, as has been observed

previously⁷, and as expected in *mtr2-21* mutants as both proteins act as a heterodimer. Notably, mutations in all Mex67 adaptor proteins, even when all Mex67 adaptor protein mutants were collectively mutated, did not affect the export of heat-shock mRNAs^{4,7} (Extended Data Fig. 1d–k). Although there is a low level of expression of heat-shock mRNAs at 37 °C (Extended Data Fig. 1g), this temperature does not induce the full heat-shock response, including an mRNA export block, which is the case at a temperature of 42 °C. For this reason we performed the heat stress analyses at 42 °C.

Genome-wide microarray and RNA sequencing (RNA-seq) experiments analysing bound RNAs after co-immunoprecipitation (co-IP) with Npl3 or Mex67 confirmed the dissociation of Npl3 and Mex67 from mRNAs during heat stress (Extended Data Fig. 2a–f and Supplementary Information). Moreover, compared to Npl3, Mex67 binds strongly to newly synthesized heat-shock transcripts, including those from heat-shock promoter element (HSE)-containing genes⁸ (Fig. 1c, Extended Data Fig. 3a, b and Supplementary Information). Quantitative reverse-transcriptase–PCR (qRT–PCR) experiments confirmed this finding for selected transcripts (Extended Data Fig. 3c).

These results suggest that bulk mRNAs are retained in the nucleus during heat shock through the dissociation of adaptor protein–Mex67 complexes, whereas freely available Mex67 selectively exports heat-shock mRNAs. For this to happen, Mex67 could bind either directly to heat-shock mRNAs or to an unknown stress-specific adaptor protein. The direct binding of Mex67 to 5S rRNA has been demonstrated previously⁹. To investigate whether Mex67 can also bind to mRNAs, we performed *in vitro* binding studies of recombinant glutathione S-transferase (GST)-tagged Mex67–Mtr2 or Npl3 with purified total RNA and found strong mRNA binding. This means that Mex67 can bind directly to mRNAs and does not discriminate between regular and stress-specific transcripts (Fig. 2a and Extended Data Fig. 4a).

The binding of rRNA to Mex67 requires the loop domain of the protein⁹ (Extended Data Fig. 4c), which also allows Mex67–Mtr2 to bind mRNAs as shown by *in vitro* RNA-binding studies using stably expressed loop-domain mutants (Fig. 2b and Extended Data Fig. 4b, d). Notably, *in vivo* these mutants also affect heat-shock mRNA export (Fig. 2c and Extended Data Fig. 4e–g).

To determine whether this loop domain also forms the binding platform for the adaptor proteins under normal conditions, we conducted co-IP experiments, assaying binding of Npl3 to the Mex67 wild-type protein and the loop-domain mutants. We found strong binding of Npl3 to wild-type Mex67, but not to the loop mutants (Fig. 2d and Extended Data Fig. 4h). A similar result was also obtained *in vitro* using purified recombinant proteins (Extended Data Fig. 4i, j), suggesting that the binding of both RNA and adaptor proteins to Mex67 requires the same loop domain. In fact, *in vitro* competition assays show that purified yeast RNA, but not DNA, disrupts a pre-formed Mex67–Npl3 complex (Fig. 2e and Extended Data Fig. 4k).

¹Abteilung für Molekulare Genetik, Institut für Mikrobiologie und Genetik, Göttinger Zentrum für Molekulare Biowissenschaften, Georg-August Universität Göttingen, Göttingen, Germany.

²Transkriptomanalyselabor, Institut für Entwicklungsbiologie, Georg-August Universität Göttingen, Göttingen, Germany.

*These authors contributed equally to this work

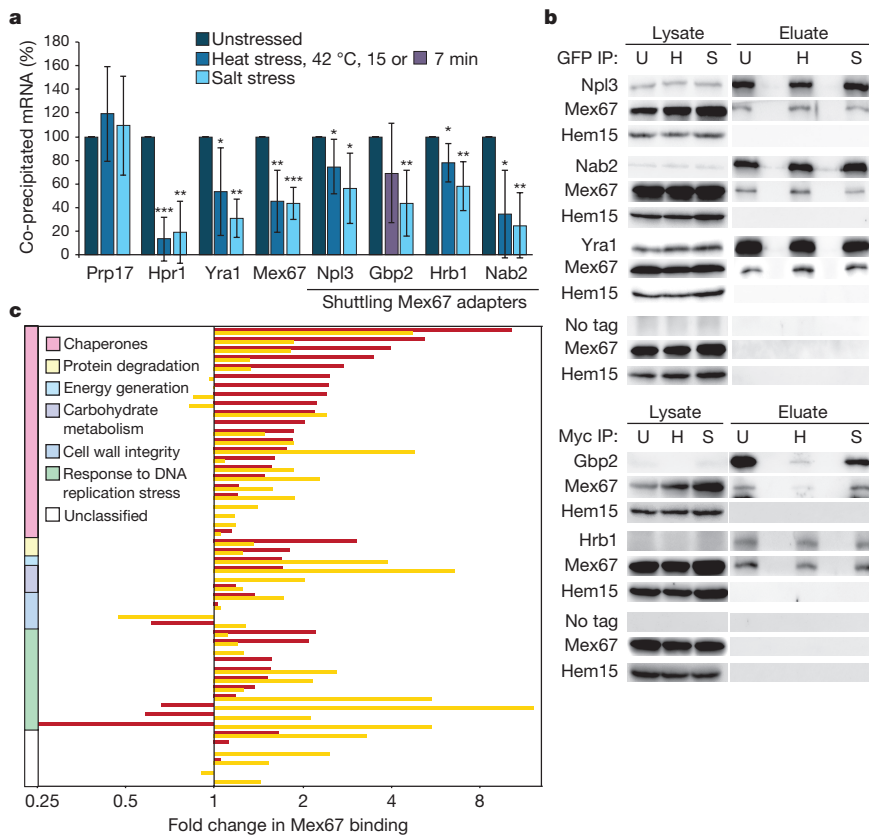


Figure 1 | Stress induces the dissociation of Mex67 and its adaptor proteins from mRNAs and Mex67-mediated export of heat-shock mRNAs. **a**, The depicted proteins were immunoprecipitated from cells treated as indicated. Co-precipitated RNA was purified and the amount of poly(A)⁺ RNA was determined by northern blotting. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$; two-tailed, two-sample unequal variance test, $n \geq 3$. **b**, Mex67 adaptor proteins dissociate from mRNA, but not from Mex67. Western blot analysis of precipitated adaptor proteins and co-precipitated Mex67 in the presence of RNase is shown. Hem15 served as negative control. No-tag controls were probed with anti-Myc or anti-GFP antibodies. H, heat stress; S, salt stress U, unstressed; IP, immunoprecipitation. **c**, RIP and sequencing (RIP-seq) (red) and microarray (yellow) studies reveal the increased binding of Mex67 to Hsf1 target genes, shown as the fold change in the binding of Mex67 to Hsf1-target gene transcripts relative to Npl3 upon shift from 25 °C to 42 °C.

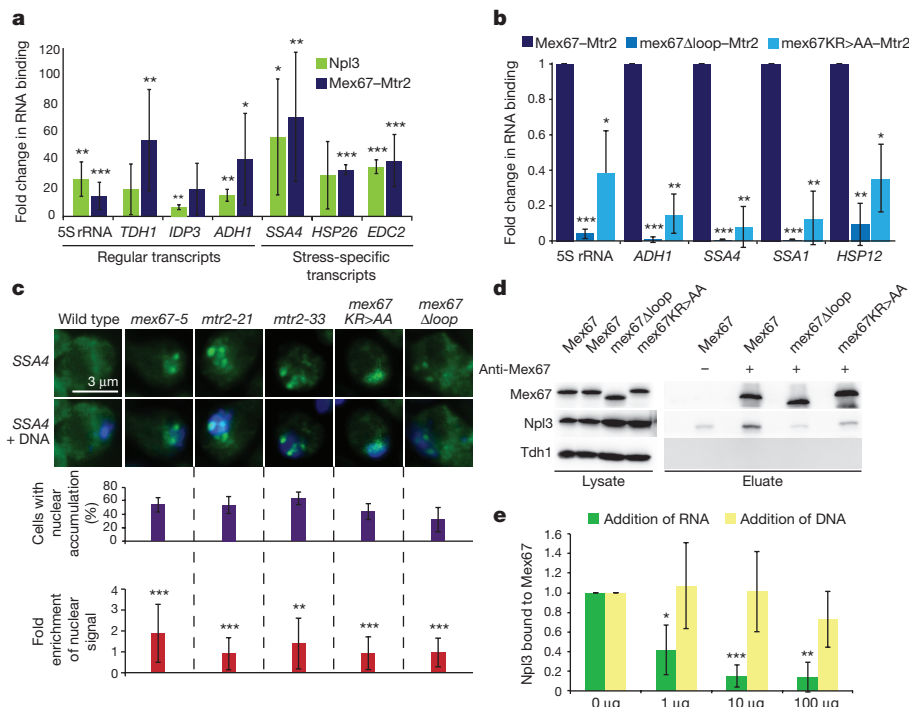


Figure 2 | Mex67 directly binds and exports heat-shock mRNAs using the same domain as it uses to interact with the adaptor protein Npl3. **a**, Mex67 binds to coding RNAs *in vitro*. Recombinant GST-Mex67/His₆-Mtr2, GST-Npl3 or a negative control GST/His₆-Mtr2, were incubated with total yeast RNA isolated from heat-stressed cells. After protein immunoprecipitation, the amount of bound mRNA was determined by qRT-PCR. Fold change is measured against binding of GST alone. **b**, Mex67 mutants with defective loop domains fail to bind mRNA *in vitro*. Purified recombinant full-length Mex67-His₆-Mtr2, indicated mutant proteins and His₆-Mtr2 (negative control) were incubated with total yeast RNA and re-purified via Ni-NTA-agarose. Bound transcript amounts

were determined by qRT-PCR. Data are normalized to the change in RNA binding of His₆-Mtr2. **c**, FISH experiments (top) and quantifications (bottom) reveal export defects of the heat-shock mRNA *SSA4* (green) in *mex67* loop-domain mutants after 30 min at 42 °C, >20 cells. **d**, The RNA-binding domain of Mex67 is required for interaction with Npl3 *in vivo*. Western blot analysis of Mex67 co-immunoprecipitation is shown. **e**, Increasing amounts of RNA, but not DNA, significantly reduce interaction of Npl3 and Mex67 in *in vitro* competition assays. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$; two-tailed, two-sample unequal variance test two-tailed, two-sample unequal variance test, $n \geq 3$ for all panels.

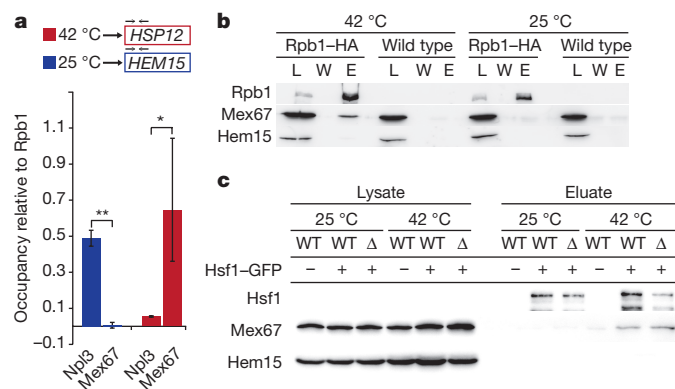


Figure 3 | Stress induces direct recruitment of Mex67 to promoters of heat-shock genes. **a**, ChIP assays reveal an increase in the binding of Mex67 to heat-shock transcripts. ChIPs with Npl3 and Mex67 in relation to Rpb1 are shown. **b**, Increased interaction of Mex67 with RNA polymerase II during heat stress. Western blot analysis of Rpb1 co-IPs detects Mex67. E, eluate; L, lysate; W, wash. **c**, Hsf1 interacts with Mex67 at 42 °C. Western blot analysis of Hsf1 co-IPs detecting Mex67 are shown. WT, wild type; Δ , *mft1* Δ . ** $P < 0.01$, * $P < 0.05$; two-tailed, two-sample equal variance test, $n \geq 3$ (**a**).

During heat stress, Mex67 binds to heat-shock transcripts that are expressed by genes that contain HSEs in their promoters (Fig. 1c). These elements are bound by the highly conserved heat-stress-specific transcriptional activator Hsf1. Hsf1 can bypass the requirements for a number of critical general transcription factors, including the C-terminal domain of RNA polymerase II (ref. 10). Instead, Hsf1 itself recruits important factors such as the Mediator transcriptional co-activator complex¹¹. Chromatin immunoprecipitation (ChIP) showed that Npl3, but not Mex67, binds to *HEM15* at 25 °C, whereas Mex67, but not Npl3, is recruited to *HSP12* at 42 °C (Fig. 3a). Consistent with this, Mex67 does not interact with the largest RNA polymerase II subunit (Rpb1) under regular conditions, but does during heat stress (Fig. 3b and Extended Data Fig. 4l). It may be that Mex67 is loaded onto the transcription machinery via Hsf1, as they interact during heat stress (Fig. 3c). Although the THO complex was shown to be required for RNA export under heat stress^{7,12}, it does not influence the interaction between Hsf1 and Mex67, as is seen in the THO complex mutant *mft1* Δ .

The direct loading of Mex67 onto heat-shock mRNAs, without the need for adaptor proteins, raises two questions. It is unclear why such a direct export mode was not established for all mRNAs, and why cells retained adaptor proteins in evolution if it is possible to export mRNA without them. In fact, the shuttling mRNA adaptor proteins have been suggested to form part of a nuclear mRNA surveillance system that prevents the nuclear export of incorrect, possibly improperly processed or assembled, ribonucleoproteins (RNPs). The principle components of this surveillance system include the TRAMP complex (Trf4/Trf5, Air1/Air2, Mtr4), which tags incorrect transcripts with short poly(A) tails, marking them for subsequent degradation by the Rps6-containing exosome¹³. Additionally, Mlp1 interacts with adaptor proteins and acts as a final gatekeeper at the nuclear pore complex (NPC) by retaining immature transcripts in the nucleus^{14–19}.

Adaptor proteins are loaded onto mRNAs at different processing steps: Npl3 associates close to the 5' cap^{20,21}, Gbp2 and Hrb1 associate during splicing^{17,20,22} and Nab2 associates preferentially at the 3' end^{20,23}. Therefore, the adaptor proteins might control individual maturation steps, marking mRNAs as export-competent via binding of Mex67 to allow subsequent Mlp1-controlled nuclear export¹⁷. To visualize their function in quality control, we utilized a FISH assay that monitors bulk mRNAs¹⁷. While mRNAs are evenly distributed in wild-type cells or in cells with *GBP2*, *HRB1* or *NPL3* deletions, the *rrp6* Δ mutant accumulates RNAs in the nucleus that are usually degraded (Fig. 4a and Extended Data Fig. 5a–d). As has been demonstrated, the

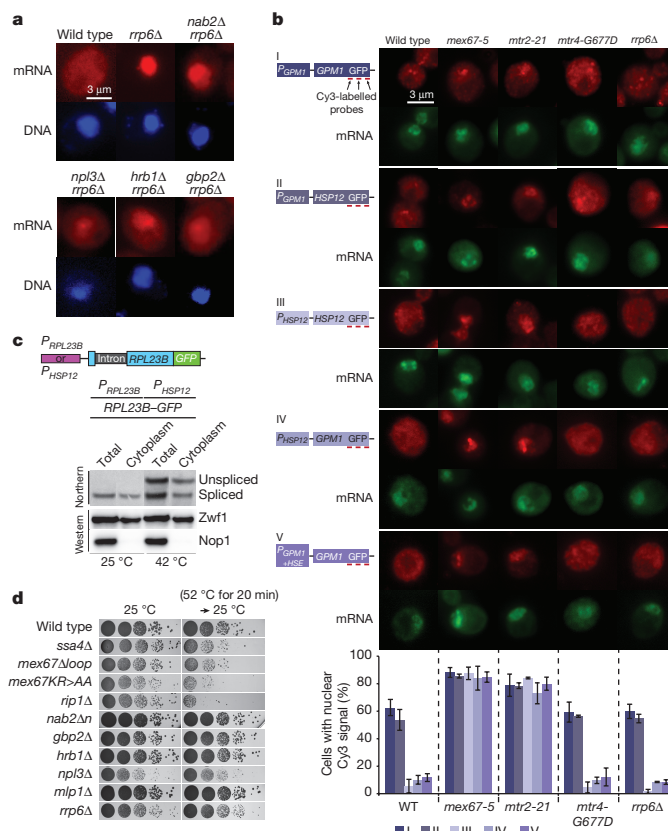


Figure 4 | Stress-specific mRNAs are not quality controlled. **a**, FISH experiments with a Cy3-labelled oligo d(T)₅₀ probe (red) show leakage at 30 °C for 3 h of mRNAs, that would ordinarily be degraded, into the cytoplasm in the absence of Mex67 adaptor proteins. **b**, Promoter-dependent accumulation of transcripts in quality-control-factor mutants. Strains carrying plasmids with indicated gene constructs were analysed in FISH experiments at 42 °C for 30 min (top) and quantified (bottom) with oligonucleotides targeting GFP mRNA (red) and poly(A)⁺ RNA (green) ($n \geq 3$, >20 cells). **c**, Incorrect transcripts are not retained in the nucleus upon heat stress. Northern blot analysis of cytoplasmic fractionation experiments detecting intron-containing *RPL23B* transcripts, expressed from the indicated promoters. **d**, mRNA quality-control factors are dispensable for the heat-shock response. Serial dilutions of severely heat-stressed (right; 52 °C for 20 min) strains are shown on full medium plates.

simultaneous deletion of *GBP2* and *HRB1* alleviates this phenotype¹⁷, as do the double mutants *npl3* Δ *rrp6* Δ and *nab2* Δ *rrp6* Δ , allowing poly(A)⁺ RNA leakage to the cytoplasm. Expression of *mex67-5* in any adaptor protein Δ *rrp6* Δ double mutant interrupted leakage, indicating that the irregular nuclear escape of false transcripts depends on functional Mex67 (Extended Data Fig. 5e, f). Whereas deletion of the adaptor proteins caused enhanced cytoplasmic mRNA leakage, the opposite was observed when intact or mutated quality control factors were overexpressed, which is toxic and leads to nuclear mRNA retention^{15,24} (Extended Data Fig. 5g–j). These results uncover an important function of Mex67 adaptor proteins in nuclear mRNA quality control. It is likely that their role is to monitor mRNA processing and recruit the degradation machinery instead of Mex67, in case processing events take too long, as has previously been shown for Gbp2 and Hrb1 (ref. 17).

As Mex67 can export heat-shock transcripts without the need for adaptor proteins, it is conceivable that these transcripts undergo no quality control. This is of particular relevance, because the RNA degradation machinery is functional during heat stress and degrades heat-shock RNAs under certain conditions, for example, when these transcripts are artificially trapped in the nucleus^{25,26}. In this situation, it is possible that the transcripts are not being properly covered by

proteins and are therefore not protected against degradation. Thus, by preventing adaptor proteins from binding to heat-shock mRNAs, their potential quality-control-mediated degradation would be circumvented. Our *in vivo* leakage assay clearly showed that, in contrast to regular mRNAs, heat-shock mRNAs do not accumulate in the nuclei of *rrp6* or *mtr4* mutants at 42 °C (Extended Data Fig. 6a–c), suggesting that stress-specific transcripts do not undergo adaptor-protein-mediated quality control. During stress, the quality control factors Nab2 and Mlp1 are eliminated from the NPCs and accumulate in nuclear foci³, which probably supports uncontrolled mRNA export.

It is possible that omitting quality control of heat-shock transcripts might speed up heat-shock protein expression, increasing the cell's chance of survival. However, this raises the question of how the cell knows to send heat-shock transcripts to 'fast-track export'. This is likely to be determined by the gene promoter, because, as has been shown²⁷, HSEs are required for translation during stress and Hsf1 recruits Mex67 (Fig. 3). Promoter-swap experiments show that *GPM1* or *CYC1* transcripts, expressed under their own promoter, strongly accumulate in *mtr4* and *rrp6* mutants. This is not the case when these genes are expressed under the control of the *HSP12* promoter or their own promoters that have been modified to contain artificial HSEs, suggesting that these RNAs are not quality controlled (Fig. 4b and Extended Data Fig. 7a–d). Vice versa, the *HSP12* transcript undergoes quality control when it is expressed under the control of the *GPM1* promoter, suggesting that the fate of undergoing quality control is encrypted in the promoter.

To demonstrate that heat-stress-responsive transcripts bypass quality control, we expressed the intron-containing housekeeping mRNA *RPL23B* under the control of the *HSP12* promoter and verified the export of unprocessed transcripts to the cytoplasm (Fig. 4c and Extended Data Fig. 7e). The skipping of quality control during heat stress on a cellular level is reflected in a recovery assay, in which the rate of cell survival is an indication of participation in the stress response. Whereas disturbances in mRNA export by, for example, mutations in *MEX67* or the nucleoporin *RIP1/NUP42* drastically affect heat-shock survival rate, the absence of quality control factors does not (Fig. 4d).

Together, our data indicate that stress precludes adaptor-protein-mediated quality control and export of bulk mRNA. Instead, it induces the direct Mex67-mediated export of heat-shock mRNA without quality control (Extended Data Fig. 7f). Thus, at the cost of accuracy, cells use a simplified mRNA export route for heat-shock transcripts, guaranteeing their rapid translation and enabling cell survival under extreme conditions.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 April 2015; accepted 24 October 2016.

Published online 12 December 2016.

1. Bond, U. Stressed out! Effects of environmental stress on mRNA metabolism. *FEMS Yeast Res.* **6**, 160–170 (2006).
2. Tutucci, E. & Stutz, F. Keeping mRNPs in check during assembly and nuclear export. *Nat. Rev. Mol. Cell Biol.* **12**, 377–384 (2011).
3. Carmody, S. R., Tran, E. J., Apponi, L. H., Corbett, A. H. & Wente, S. R. The mitogen-activated protein kinase Sit2 regulates nuclear retention of non-heat shock mRNAs during heat shock-induced stress. *Mol. Cell. Biol.* **30**, 5168–5179 (2010).
4. Krebber, H., Taura, T., Lee, M. S. & Silver, P. A. Uncoupling of the hnRNP Npl3p from mRNAs during the stress-induced block in mRNA export. *Genes Dev.* **13**, 1994–2004 (1999).
5. Wallace, E. W. *et al.* Reversible, specific, active aggregates of endogenous proteins assemble upon heat stress. *Cell* **162**, 1286–1298 (2015).

6. Mitchell, S. F., Jain, S., She, M. & Parker, R. Global analysis of yeast mRNPs. *Nat. Struct. Mol. Biol.* **20**, 127–133 (2013).
7. Rollenhagen, C., Hodge, C. A. & Cole, C. N. Following temperature stress, export of heat shock mRNA occurs efficiently in cells with mutations in genes normally important for mRNA export. *Eukaryot. Cell* **6**, 505–513 (2007).
8. Yamamoto, A., Mizukami, Y. & Sakurai, H. Identification of a novel class of target genes and a novel type of binding sequence of heat shock transcription factor in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **280**, 11911–11919 (2005).
9. Yao, W. *et al.* Nuclear export of ribosomal 60S subunits by the general mRNA export receptor Mex67-Mtr2. *Mol. Cell* **26**, 51–62 (2007).
10. Singh, H. *et al.* A functional module of yeast mediator that governs the dynamic range of heat-shock gene expression. *Genetics* **172**, 2169–2184 (2006).
11. Kim, S. & Gross, D. S. Mediator recruitment to heat shock genes requires dual Hsf1 activation domains and mediator tail subunits Med15 and Med16. *J. Biol. Chem.* **288**, 12197–12213 (2013).
12. Strässer, K. *et al.* TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417**, 304–308 (2002).
13. Doma, M. K. & Parker, R. RNA quality control in eukaryotes. *Cell* **131**, 660–668 (2007).
14. Burkard, K. T. & Butler, J. S. A nuclear 3'-5' exonuclease involved in mRNA degradation interacts with Poly(A) polymerase and the hnRNA protein Npl3p. *Mol. Cell. Biol.* **20**, 604–616 (2000).
15. Galy, V. *et al.* Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1. *Cell* **116**, 63–73 (2004).
16. Green, D. M., Johnson, C. P., Hagan, H. & Corbett, A. H. The C-terminal domain of myosin-like protein 1 (Mlp1p) is a docking site for heterogeneous nuclear ribonucleoproteins that are required for mRNA export. *Proc. Natl Acad. Sci. USA* **100**, 1010–1015 (2003).
17. Hackmann, A. *et al.* Quality control of spliced mRNAs requires the shuttling SR proteins Gbp2 and Hrb1. *Nat. Commun.* **5**, 3123 (2014).
18. Inoue, K., Mizuno, T., Wada, K. & Hagiwara, M. Novel RING finger proteins, Air1p and Air2p, interact with Hmt1p and inhibit the arginine methylation of Npl3p. *J. Biol. Chem.* **275**, 32793–32799 (2000).
19. Schmid, M. *et al.* Rrp6p controls mRNA poly(A) tail length and its decoration with poly(A) binding proteins. *Mol. Cell* **47**, 267–280 (2012).
20. Baejen, C. *et al.* Transcriptome maps of mRNP biogenesis factors define pre-mRNA recognition. *Mol. Cell* **55**, 745–757 (2014).
21. Shen, E. C., Stage-Zimmermann, T., Chui, P. & Silver, P. A. 7The yeast mRNA-binding protein Npl3p interacts with the cap-binding complex. *J. Biol. Chem.* **275**, 23718–23724 (2000).
22. Tuck, A. C. & Tollervey, D. A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. *Cell* **154**, 996–1009 (2013).
23. Schmid, M. *et al.* The nuclear polyA-binding protein Nab2p is essential for mRNA production. *Cell Reports* **12**, 128–139 (2015).
24. Windgassen, M. & Krebber, H. Identification of Gbp2 as a novel poly(A)⁺ RNA-binding protein involved in the cytoplasmic delivery of messenger RNAs in yeast. *EMBO Rep.* **4**, 278–283 (2003).
25. Hilleren, P., McCarthy, T., Rosbash, M., Parker, R. & Jensen, T. H. Quality control of mRNA 3'-end processing is linked to the nuclear exosome. *Nature* **413**, 538–542 (2001).
26. Thomsen, R., Libri, D., Boulay, J., Rosbash, M. & Jensen, T. H. Localization of nuclear retained mRNAs in *Saccharomyces cerevisiae*. *RNA* **9**, 1049–1057 (2003).
27. Zid, B. M. & O'Shea, E. K. Promoter sequences direct cytoplasmic localization and translation of mRNAs during starvation in yeast. *Nature* **514**, 117–121 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank H. Bastians for advice, L. Oldehaver for technical assistance, W. Kramer for discussion and R. Lill, P. A. Silver, C. Dargemont and E. Hurt for providing plasmids, strains or antibodies. This work was funded by grants from the Deutsche Forschungsgemeinschaft and the SFB860 to H.K.

Author Contributions Experiments were designed and data interpreted by all authors; experiments were performed by G.Z. (Figs 1b, c, 2d, e, 3c, 4b–d), A.H. (Figs 1c, 3a, b, 4a), L.B. (Figs 1a, 2a–c) D.B. (Fig. 1c) and T.L. and G.S. (Fig. 1c). The manuscript was written by H.K.; all authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.K. (heike.krebber@biologie.uni-goettingen.de).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Yeast strains, plasmids and oligonucleotides. All yeast strains used in this study are listed in Extended Data Table 1, oligonucleotides in Extended Data Table 2 and plasmids in Extended Data Table 3. Plasmids and yeast strains were generated by conventional methods.

Co-immunoprecipitation experiments. The experiments were essentially performed as published previously⁷. All yeast strains were grown to log phase ($2-3 \times 10^7$ cells/ml). For the experiment shown in Fig. 1b, the cells were split into three portions: one served as an unstressed control and the cells of the two other portions were either incubated for 15 min at 42 °C to apply heat stress or treated with 1 M NaCl for 1 h at 25 °C for salt stress. Cells for the precipitation shown in Fig. 2d were not shifted. For co-IP analyses shown in Fig. 3b, c, cells were split into two portions: one served as an unstressed control (25 °C) and the other was incubated for 15 min at 42 °C. Afterwards, the cells were collected immediately and lysed in immunoprecipitation buffer (1 × PBS, 3 mM KCl, 2.5 mM MgCl₂, 0.5% Triton X-100, 200 µg/ml RNase A, vanadyl phosphatase inhibitors and protease inhibitors from Roche). The supernatant was incubated for 3–4 h at 4 °C (Figs 1b, 2d) or for 20 min, at either 25 °C (for the unstressed cell lysate) or 37 °C (for the heat-stressed lysate) (Fig. 3b, c) with protein G sepharose beads (Amersham Biosciences) bound to Myc (9E10)-antibody (Santa Cruz), HA-antibody (Santa Cruz) or Mex67 antiserum (own) or with GFP-Trap_A beads (Chromotek). The matrix was washed six times with immunoprecipitation buffer and proteins were detected by western blot analyses with the indicated antibodies (GFP (Pierce) 1:5,000; Myc (9E10) (Santa Cruz) 1:1,000; Tdh1 (Pierce) 1:5,000; Hem15 (R. Lill) 1:7,000; Mex67 (C. Dargemont) 1:2,000; Mex67 (rabbit, serum) 1:50,000; HA (Santa Cruz) 1:1,000, Rps3 (rabbit, own serum) 1:750). Signals were detected with the Fusion SL system (PiqLab). Intensities were quantified using the Bio1D software.

RIP experiments. The experiments were essentially carried out as described previously¹⁷. All yeast strains were grown to mid-log phase ($2-3 \times 10^7$ cells/ml). For the experiment shown in Fig. 1a and Extended Data Fig. 1c, cells were split into three portions: one cell portion served as an unstressed control, one portion was incubated for 15 min (7 min for Gbp2) at 42 °C to apply heat stress and the final portion was treated with 1 M NaCl for 1 h at 25 °C for salt stress. Afterwards the cells were immediately collected and lysed in RIP buffer (25 mM Tris HCl pH 7.5, 100 mM KCl, 0.2% (v/v) Triton X-100, 0.2 mM PMSF, 5 mM DTT, 10 U RiboLock RNase Inhibitor (Thermo Scientific) and protease inhibitor (Roche)). The supernatant was incubated for 3–4 h at 4 °C with protein G sepharose beads (Amersham Biosciences) bound to Myc (9E10)-antibody (Santa Cruz) or with GFP-Trap_A beads (Chromotek). The matrix was washed six times with RIP buffer and split into two portions after the last washing step. Proteins were detected by western blot. RNA was extracted using phenol/chloroform and further purified using an RNA purification kit (Macherey & Nagel). All of the purified RNA was used for subsequent dot-blot experiments.

For Fig. 1c and Extended Data Figs 2, 3 (microarray and RNA-seq) yeast cells were grown at 25 °C in yeast extract peptone dextrose (YPD) medium to log phase. The cells were collected and resuspended in 50 ml YPD medium and incubated at either 25 °C or 42 °C in a Petri dish for 30 min. Subsequently, cells were irradiated with UV (254 nm, 120,000 MJ/cm²) to induce cross-linking. After centrifugation, cell pellets were washed in 1 ml RIP buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.2% (v/v) Triton X-100, 2 mM DTT, 10 U RiboLock RNase Inhibitor (Thermo Scientific)). One volume of the pellet was mixed with 1.5 volumes of RIP buffer, containing EDTA-free protease inhibitor cocktail (Roche) and 1 volume of glass beads. Cells were lysed by vigorous mixing for 30 s at 4 m/s using the FastPrep-24 machine (MP Biomedicals). Co-IP experiments were performed at 4 °C for 4 h by incubating the lysates with protein G sepharose beads bound to monoclonal Myc (9E10) or GFP-Trap_A beads. Afterwards beads were washed five times with RIP buffer and treated with 20 µg proteinase K at 55 °C for 20 min to remove RNA-bound proteins. Subsequently, eluates were purified via trizol–chloroform (Ambion RNA, Life Technologies) extraction. Contaminating DNA was removed with the TURBO DNA-free kit (Ambion RNA by Life technologies). The purified RNA was reverse transcribed with Maxima reverse transcriptase (Thermo Scientific) for subsequent qRT-PCR analyses.

Dot-blot analyses. For dot blot experiments (Fig. 1a and Extended Data Fig. 1a, c) RNA was spotted onto a Hybond N⁺ nylon membrane (GE Healthcare) and UV cross-linked (254 nm, 120,000 µJ/cm²). The membrane was incubated at 80 °C for 2 h, pre-hybridized for 1 h in hybridization buffer (0.5 M sodium phosphate buffer, pH 7.5, 7% SDS, 1 mM EDTA) and hybridized overnight at 42 °C in hybridization buffer with a digoxigenin (DIG)-labelled oligo d(T)₅₀ probe. The membrane was washed once with 2 × SSC buffer, 0.1% (w/v) SDS; once with 1 × SSC buffer, 0.1% (w/v) SDS for 15 min at room temperature; and twice with 0.5 × SSC buffer,

0.1% (w/v) SDS for 15 min at 42 °C. For detection of the signal, the manufacturer's instructions (Roche) were followed. Signal intensities were quantified using the Fusion camera (PiqLab) and Fiji software, and compared to the signal of the unstressed samples, always in relation to the precipitated protein.

FISH experiments. The experiments were essentially carried out as described previously¹⁷. RNA probes were synthesized by *in vitro* transcription, using the T7 RNA polymerase (Thermo Scientific) and labelled with DIG-UTP using an RNA labelling mix (Roche) or with Cy3-labelled oligonucleotides (Sigma), which are listed in Extended Data Table 2. To detect poly(A)⁺ RNA a Cy3- or Atto488-labelled oligo d(T)₅₀ probe (Sigma) was used. Cells were grown to log phase before being subjected to heat stress at 42 °C for 30 min or 1 h. For leakage analyses, cells were shifted to 30 °C (Fig. 4a and Extended Data Fig. 5a, f, h) or to 37 °C (Extended Data Fig. 5c, j) for 3 h. For visualization of mRNAs under control of the *GPM1* promoter, cells were grown in 1% glucose (in the respective medium). Glucose concentration was adjusted to 4% at 15 min before heat stress. Samples were fixed by adding formaldehyde, giving a final concentration of 4%. Cells were spheroplasted by adding zymolyase, permeabilized in 0.1 M potassium phosphate buffer pH 6.5 with 1.2 M sorbitol and 0.5% Triton X-100, pre-hybridized with Hybmix (50% deionized formamide, 5 × SSC buffer, 1 × Denhardt's, 500 µg/ml tRNA, 500 µg/ml salmon sperm DNA, 50 µg/ml heparin, 2.5 mM EDTA pH 8.0, 0.1% Tween 20, 10% dextran sulfate) for 1 h on a polylysine coated slide at 37 °C and hybridized in Hybmix with the specific probe overnight at 37 °C. After hybridization, cells were washed with 2 × SSC buffer and 1 × SSC buffer at room temperature, each for 1 h and with 0.5 × SSC buffer at 37 °C and room temperature, each for 30 min. For detection of DIG probes, the cells were treated with blocking buffer containing 5% heat-inactivated FBS for 1 h and incubated with sheep anti-digoxigenin Fab-FITC antibody (Roche) overnight at 4 °C or at room temperature for 4 h. DNA was stained with Hoechst 33342 (Sigma). Microscopy studies were performed with a Leica AF6000 microscope and pictures were obtained using the LEICA DFC360FX camera and the LAS AF 2.7.3.9 software (Leica) and quantified using Fiji software.

Microarray and RNA-seq analyses. A detailed description is given in the Supplementary Information.

In vitro binding studies. To purify yeast RNAs bound to recombinant Mex67 and Mtr2, *Escherichia coli* BL21 DE3 cells carrying pET8c-His₆-MTR2 (pHK1279) and either pGEX4T-1-GST-MEX67 (pHK442) or pGEX4T-1-GST-NPL3 (pHK1276), or, as a negative control, pGEX4T-1-GST (pHK439). pET8c-His₆-MTR2 (pHK1279) cells were grown for two days at 16 °C with 2 mM isopropyl-β-D-thiogalactopyranosid (IPTG). Cells were collected and sonicated in binding buffer (5% (v/v) glycerol, 100 mM NaCl, 2 mM MgCl₂, 20 mM HEPES, 0.14% (v/v) 2-β-mercaptoethanol and protease inhibitor from Roche). Protein complexes were purified via affinity purification with Glutathione Sepharose 4B (GE Healthcare). The sepharose (20 µl of 50% slurry) was incubated with bacterial lysate containing the respective recombinant proteins and purified total yeast RNA from stressed and unstressed cells (50 µg/sample) for 1 h at 4 °C. The sepharose was washed six times with binding buffer and split into two portions, one of which contained 10% of the beads to control the pull-down efficiency of the western blot analysis. The other fraction was treated with proteinase K at 37 °C for 30 min. RNA was extracted with phenol/chloroform and further purified with an RNA purification kit (Macherey & Nagel). cDNA was prepared with Maxima Reverse Transcriptase (Thermo Scientific) with random hexamer primers according to the manufacturer's instructions. Abundance of mRNA was determined by qRT-PCR using GoTaq 2 × master mix (Promega) and primers specific for each transcript (Extended Data Table 2). The mRNA levels were compared to the negative control (Fig. 2a).

To investigate whether the mutations in the loop domain of Mex67 (ref. 9) would affect mRNA binding (Fig. 2b), recombinant His₆-tagged Mtr2 and untagged Mex67 pHK1372 (*HIS:TEV:MTR2:MEX67*), pHK1373 (*HIS:TEV:MTR2:MEX67loopKR > AA*) and pHK1374 (*HIS:TEV:MTR2:MEX67-409-435aaK343E*) (constructs described in ref. 9) were co-expressed and purified from *E. coli* Rosetta 2 cells by affinity chromatography using Ni-NTA agarose (Macherey & Nagel). As the negative control, pET8c-His₆-MTR2 (pHK1279) was expressed and purified as described above. The purified proteins were diluted in washing buffer (30 mM HEPES pH 7.5, 100 mM NaCl, 10% (v/v) glycerol, and protease inhibitors (Roche) to a final concentration of 0.45 mg/ml. Samples were incubated with purified total yeast RNA (100 µg per sample) extracted from heat-stressed cells (30 min at 42 °C) and Ni-NTA agarose (Qiagen) for 60 min at 4 °C. The matrix was treated and divided as described above. Pull down of proteins was analysed using western blot, RNA was analysed as described above. The mRNA levels were compared to the negative control and set into relation with wild type.

Mex67 in vitro co-IP and RNA/DNA competition assay. GST-Npl3 was recombinantly expressed in *E. coli* Rosetta 2 cells carrying a pGEX4T-1-GST-NPL3 (pHK1276) plasmid. Cells were collected and bacterial lysate with app. 1 µg/µl GST-Npl3 in lysis buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 4 mM MgCl₂

10% (v/v) glycerol, 1 mM DTT, 0.1% NP-40 and protease inhibitors (Roche) was prepared. Purification of recombinant Mex67, mex67 Δ loop409–435 and mex67–loopKR > AA (constructs described in ref. 9) proteins in complex with Mtr2 has already been described in the section on *in vitro* binding studies. For the co-IP experiments 50 μ g of the respective Mex67 protein, 250 μ l GST–Npl3 bacterial lysate, 100 μ g RNase A, 550 μ l lysis buffer with protein G sepharose beads (Amersham Biosciences) loaded with anti-Mex67 antibody (rabbit, own serum) were mixed and incubated at 4 °C for 1.5 h. The matrix was washed six times with lysis buffer and proteins were detected by western blot with the indicated antibodies (anti-Mex67 (rabbit, serum) 1:50,000; anti-GST (mouse, Santa Cruz) 1:2,000). The anti-Mex67 (and not the anti-GST or anti-His) antibodies were used preferentially after generation of a highly specific antibody against Mex67.

For RNA/DNA competition analysis, co-IP was performed as above, but the matrix was washed only three times. Following this, 1 ml lysis buffer, 1 μ l Ribolock (Thermo Scientific) and the indicated amount of total yeast RNA (prepared from cells that were shifted to 42 °C for 30 min) or DNA (a 7.3-kb plasmid was digested with BanI, resulting in 6 fragments from 94 bp to 3.1 kb) were added and incubated for 1 h at 4 °C. The beads were washed four times with lysis buffer and proteins were detected as described above. Signal intensities were quantified using the Bio1D software.

ChIP experiments. For ChIP experiments, cells were grown at 25 °C to OD₆₀₀ of 0.8. Cultures were split into two equal parts and either shifted to 42 °C or further incubated at 25 °C for 20 min, crosslinked with 1% formaldehyde for 20 min and quenched with 250 mM glycine for 5 min. Cells were collected and washed three times with TBS (20 mM Tris–HCl, pH 7.5, 150 mM NaCl). Pellets were lysed with 1 \times pellet volume of glass beads and 3 \times pellet volume of ChIP lysis buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 1% Triton X-100, 0.1% sodium deoxycholate, 1 mM EDTA, 0.1% SDS, 1 mM PMSF) for 30 s at 5.5 m/s two times using the FastPrep machine. The lysates were sonicated in a water bath sonicator at 100% duty for 2.5 min to obtain a ~350-bp DNA fragment size. A total of 10% of the lysate was collected and used as an immunoprecipitation input control. Immunoprecipitation with cleared lysate was performed using anti-Myc (Santa Cruz) and anti-HA antibodies (Santa Cruz) and G-Sepharose beads (GE) or with GFP trap beads (Chromotek) for 3 h at 4 °C. Beads were washed twice with lysis buffer, one time with high-salt buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 1% TritonX-100, 0.1% sodium deoxycholate, 1 mM EDTA), deoxycholate wash buffer (10 mM Tris pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA) and twice with TE buffer pH 8.0 (10 mM Tris, 1 mM EDTA). Immunoprecipitation eluates and input samples were treated with proteinase K for 1 h at 42 °C and crosslinks were reversed by overnight incubation at 65 °C. DNA was purified via phenol/chloroform extraction. qRT–PCR analyses with input and immunoprecipitation samples were performed on a Rotor-Gene Q cycler to analyse binding at the 5'-regions of the housekeeping gene *HEM15* and the stress-regulated *HSP12* gene. As a control, primers for a non-transcribed region (NTR) of chromosome V were used and ΔC_t values were calculated from respective input and immunoprecipitation samples. Standard curves were used to determine primer efficiencies. Occupancies were calculated relative to the NTR control, according to $(E^{\Delta C_t}_{Pos} - \Delta C_{No tag}^{NTR}) / E^{\Delta C_t}_{Pos} - \Delta C_{No tag}^{GOI}$ (Pos., tagged strains; GOI, gene of interest).

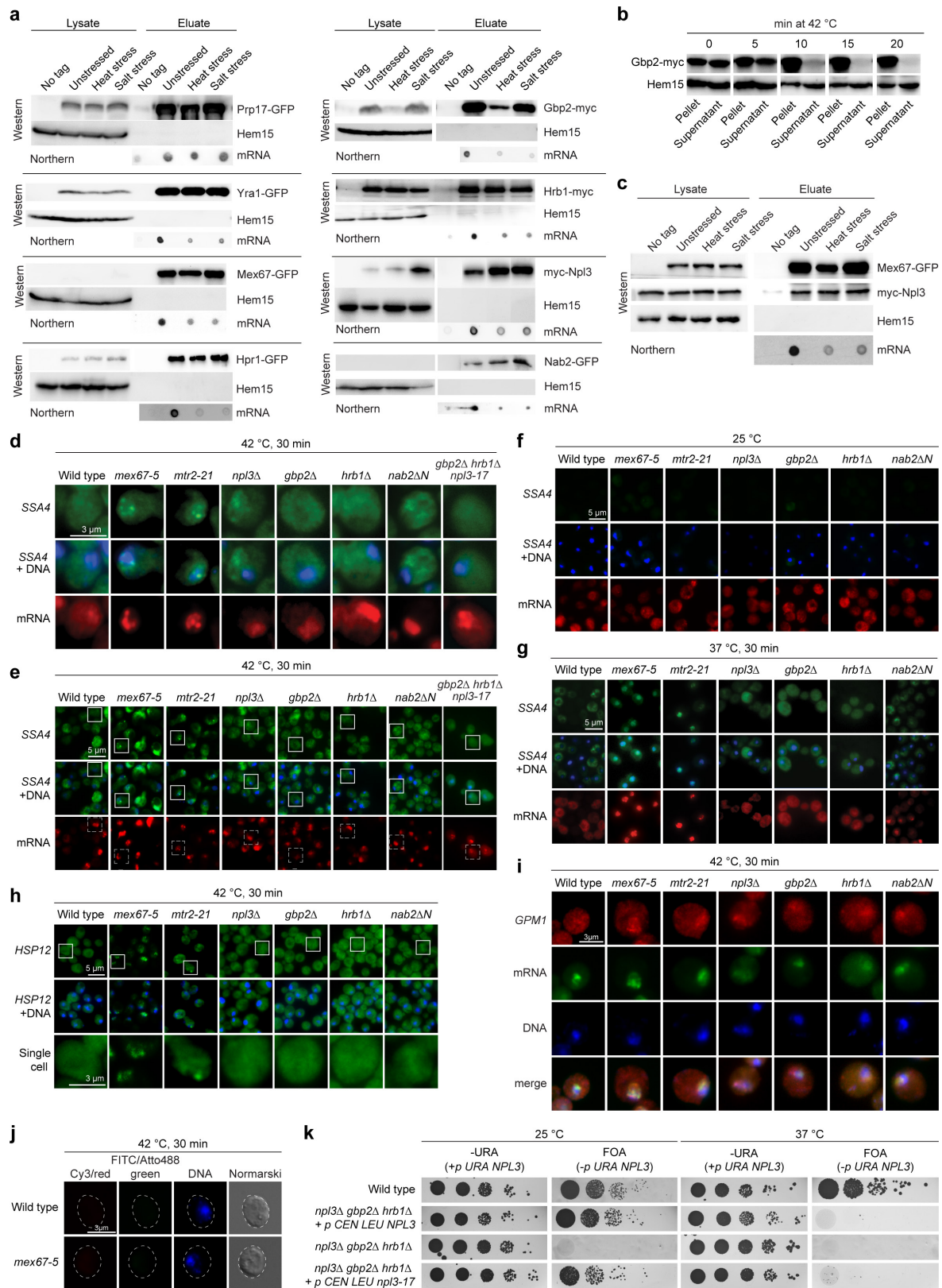
Cytoplasmic fractionation and RNA splicing analysis. For detection of unspliced mRNAs in the cytoplasm (Fig. 4c), cells were grown to log phase in selective medium. After collection, cells were washed once with 1 ml YPD/1 M sorbitol/2 mM DTT and resuspended in YPD/1 M sorbitol/1 mM DTT. Cells were spheroblasted using Zymolyase and diluted in 50 ml YPD/1 M sorbitol to recover for 30 min before shift to 42 °C for 30 min. Cells were put on ice, centrifuged at 900g for 5 min and resuspended in 500 μ l Ficoll buffer (18% Ficoll 400, 10 mM HEPES pH 6.0). Cells were lysed by addition of 1 ml buffer A (50 mM NaCl, 1 mM MgCl₂, 10 mM HEPES pH 6.0). The suspension was mixed and centrifuged at 1,500g for 15 min. The supernatant was used for cytoplasmic analyses. To verify correct fractionation of the cytoplasmic lysates, samples were analysed by western blot for the presence of the cytoplasmic Zwf1 and nucleolar Nop1. RNA was purified using a purification kit (Macherey & Nagel). Equal amounts were separated on a 1% agarose gel (1 \times MOPS, 2% formaldehyde) and blotted overnight on a Hybond–Nylon membrane. The membrane was processed as described for the dot blot experiments. For detection of the GFP-containing mRNA a DIG-labelled probe was prepared as described for FISH analyses.

Quantification. All experiments shown in this work were performed at least three times independently as biological replicates, with the exception of the genome-wide analyses. Error bars represent the standard deviation. *P* values shown in

Figs 1a, 2a, b, e and Extended Data Figs 3c, 4h, j were calculated using a two-tailed, two-sample unequal variance *t*-test. *P* values shown in Fig. 3a and Extended Data Fig. 5b c were calculated by a two-tailed, two-sample equal variance test. *P* values are indicated as follows: ****P* < 0.001, ***P* < 0.01, **P* < 0.05. For quantification of cells with displayed phenotypes (Figs 2c, 4b and Extended Data Figs 5b, c, 6a and 7d), a minimum of 20 cells was counted for each experiment. Fluorescent signal intensities were analysed using the Fiji software. The Hoechst signal indicated the nuclear area. Increased nuclear signal for Cy3 indicated nuclear accumulation of the analysed RNA and was counted as such. For intensity analyses, this area was measured for all RNA-probes on a single plane to get the intensity of the nuclear signal and this signal was set into relation with the whole-cell signal to determine the cytoplasmic signal.

Data availability. All datasets generated and analysed during the current study are available in Supplementary Information Fig. 1 and from the corresponding author upon reasonable request. All microarray and RIP-seq data that support the findings of the study have been deposited at the NCBI gene expression omnibus (GEO; www.ncbi.nlm.nih.gov/geo/) with the GEO accession numbers GSE83267 (Microarray) and GSE81542 (RNA-seq).

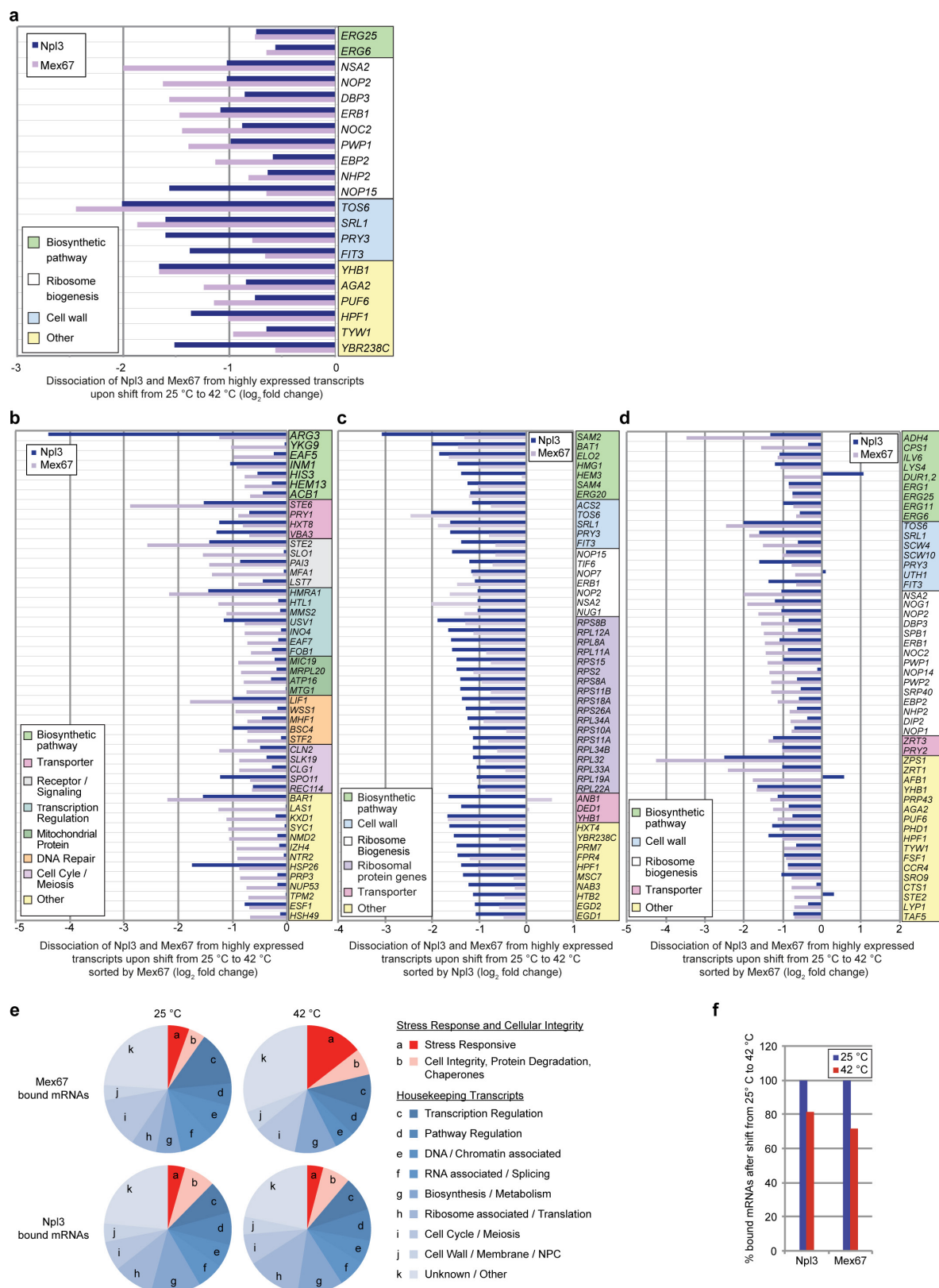
28. Winston, F., Dollard, C. & Ricupero-Hovasse, S. L. Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C. *Yeast* **11**, 53–55 (1995).
29. Gorsch, L. C., Dockendorff, T. C. & Cole, C. N. A conditional allele of the novel repeat-containing yeast nucleoporin RAT7/NUP159 causes both rapid cessation of mRNA export and reversible clustering of nuclear pore complexes. *J. Cell Biol.* **129**, 939–955 (1995).
30. Shen, E. C. *et al.* Arginine methylation facilitates the nuclear export of hnRNP proteins. *Genes Dev.* **12**, 679–691 (1998).
31. Strässer, K., Bassler, J. & Hurt, E. Binding of the Mex67p/Mtr2p heterodimer to FXFG, GLFG, and FG repeat nucleoporins is essential for nuclear mRNA export. *J. Cell Biol.* **150**, 695–706 (2000).
32. Segref, A. *et al.* Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)⁺ RNA and nuclear pores. *EMBO J.* **16**, 3256–3271 (1997).
33. Bassler, J. *et al.* Identification of a 60S preribosomal particle that is closely linked to nuclear export. *Mol. Cell* **8**, 517–529 (2001).
34. Hackmann, A., Gross, T., Baierlein, C. & Krebber, H. The mRNA export factor Npl3 mediates the nuclear export of large ribosomal subunits. *EMBO Rep.* **12**, 1024–1031 (2011). 10.1038/embor.2011.155
35. Baierlein, C. *et al.* Monosome formation during translation initiation requires the serine/arginine-rich protein Npl3. *Mol. Cell Biol.* **33**, 4811–4823 (2013). 10.1128/MCB.00873-13
36. Marfatia, K. A., Crafton, E. B., Green, D. M. & Corbett, A. H. Domain analysis of the *Saccharomyces cerevisiae* heterogeneous nuclear ribonucleoprotein, Nab2p. Dissecting the requirements for Nab2p-facilitated poly(A) RNA export. *J. Biol. Chem.* **278**, 6731–6740 (2003).
37. Taura, T., Krebber, H. & Silver, P. A. A member of the Ran-binding protein family, Yrb2p, is involved in nuclear protein export. *Proc. Natl Acad. Sci. USA* **95**, 7427–7432 (1998).
38. Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19–27 (1989).
39. Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H. & Hieter, P. Multifunctional yeast high-copy-number shuttle vectors. *Gene* **110**, 119–122 (1992).
40. Lee, M. S., Henry, M. & Silver, P. A. A protein that shuttles between the nucleus and the cytoplasm is an important mediator of RNA export. *Genes Dev.* **10**, 1233–1246 (1996).
41. Häcker, S. & Krebber, H. Differential export requirements for shuttling serine/arginine-type mRNA-binding proteins. *J. Biol. Chem.* **279**, 5049–5052 (2004).
42. Santos-Rosa, H. *et al.* Nuclear mRNA export requires complex formation between Mex67p and Mtr2p at the nuclear pores. *Mol. Cell Biol.* **18**, 6826–6838 (1998).
43. Duncan, K., Umen, J. G. & Guthrie, C. A putative ubiquitin ligase required for efficient mRNA export differentially affects hnRNP transport. *Curr. Biol.* **10**, 687–696 (2000).
44. Gilbert, W., Siebel, C. W. & Guthrie, C. Phosphorylation by Sky1p promotes Npl3p shuttling and mRNA dissociation. *RNA* **7**, 302–313 (2001).
45. Zenklusen, D., Vinciguerra, P., Strahm, Y. & Stutz, F. The yeast hnRNP-Like proteins Yra1p and Yra2p participate in mRNA export through interaction with Mex67p. *Mol. Cell Biol.* **21**, 4219–4232 (2001).
46. Conrad, N. K. *et al.* A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. *Genetics* **154**, 557–571 (2000).
47. Strambio-de-Castillia, C., Blobel, G. & Rout, M. P. Proteins connecting the nuclear pore complex with the nuclear interior. *J. Cell Biol.* **144**, 839–855 (1999).
48. Faza, M. B., Chang, Y., Occhipinti, L., Kemmler, S. & Panse, V. G. Role of Mex67-Mtr2 in the nuclear export of 40S pre-ribosomes. *PLoS Genet.* **8**, e1002915 (2012).



Extended Data Figure 1 | See next page for caption.

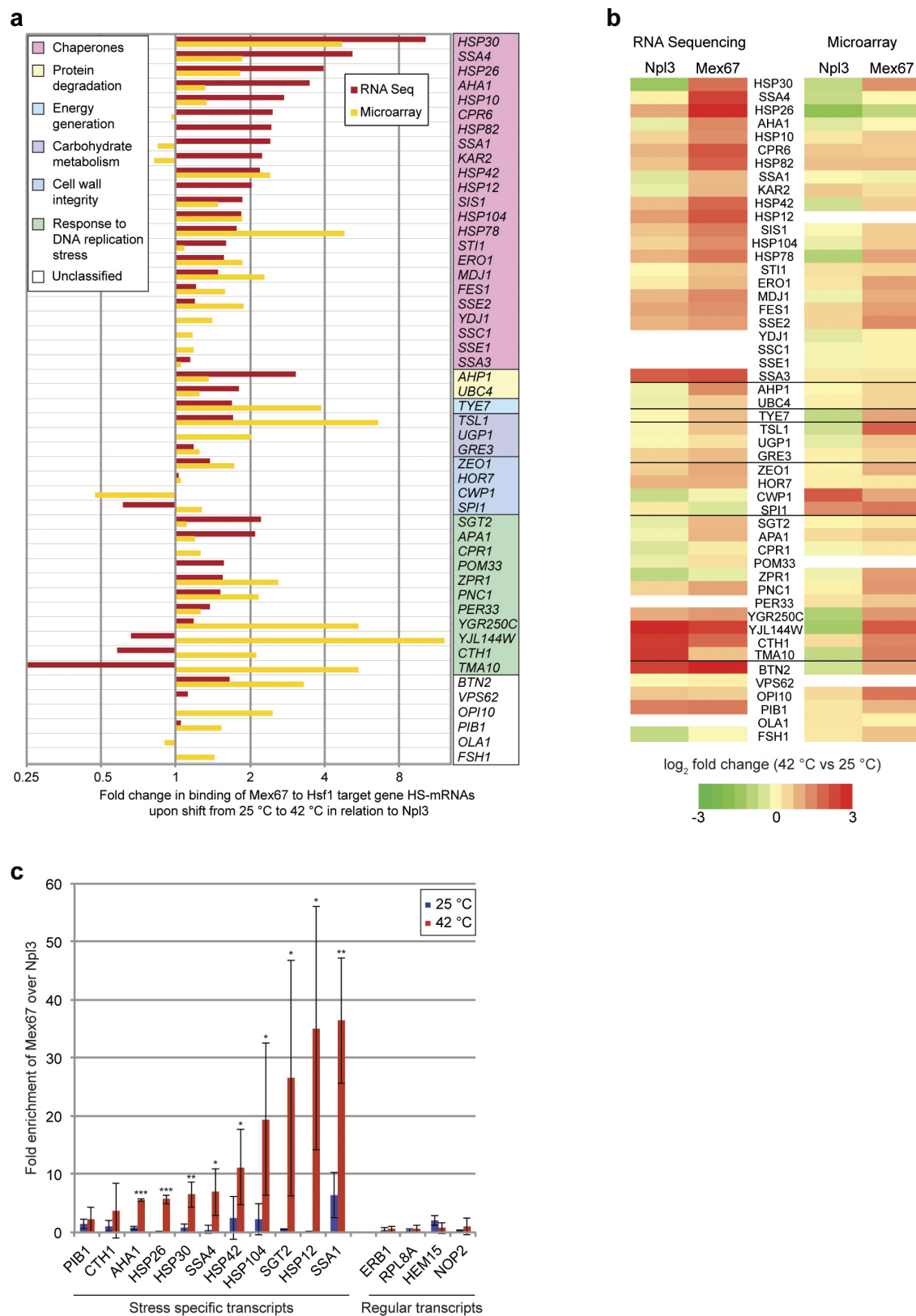
Extended Data Figure 1 | Mex67 dissociates from poly(A)⁺ RNA together with the adaptor proteins, which are dispensable for heat-shock mRNA export. **a**, Stress leads to the dissociation of Mex67 and its adaptor proteins from bulk mRNAs. Example experiments of immunoprecipitations from Fig. 1a with the indicated GFP-tagged or Myc-tagged proteins are shown in western blots. The co-precipitated poly(A)⁺RNA was detected in northern blots with a DIG-labelled oligo d(T)₅₀ probe. **b**, Gbp2 precipitates after extended heat stress. Western blot analysis for the presence of Gbp2 in the soluble supernatant fraction or the insoluble pellet fraction is shown. **c**, Mex67 dissociates from mRNA, but not from its adaptor protein. Immunoprecipitation analysis of Mex67 and the co-precipitated Npl3 is shown in western blots. Mex67-bound poly(A)⁺RNA was analysed in northern blots. **d**, FISH experiments of the *SSA4* heat-shock mRNA reveal export defects in mutants of *MEX67-MTR2*, but not the Mex67 adaptor proteins, upon a 30-min temperature shift to 42 °C. A DIG-labelled *SSA4* RNA probe (green) detects the

heat-shock mRNA in the indicated strains. DNA was stained with Hoechst (blue) and poly(A)⁺RNA with an oligo d(T)₅₀ probe (red). **e**, Overview of the experiment shown in **d**. The frame indicates the enlarged single cells in **d**. **f**, Control experiment that shows the experiment described in **d** at 25 °C, in which heat-shock RNAs are not produced in visible amounts. **g**, The experiment was performed as described in **d**, with the cells shifted to 37 °C for 30 min. **h**, The same experiment as described in **d** with a different heat-shock RNA probe against *HSP12*. **i**, The same experiment as in **d** with probes against a single mRNA, *GPM1* (Cy3, red) and poly(A)⁺ (Atto488, green). **j**, Control experiment that shows the experiment described in **i** without the DIG-labelled probe (green) or a Cy3 probe (red). The circle indicates the contour of the cells. **k**, Deletion of all three SR-protein genes is lethal to cells. The indicated strains were spotted in serial dilution onto agar plates selecting either for a covering plasmid (plates lacking uracil, –URA) or for the loss of this plasmid (5-fluoroorotic acid plates, FOA).



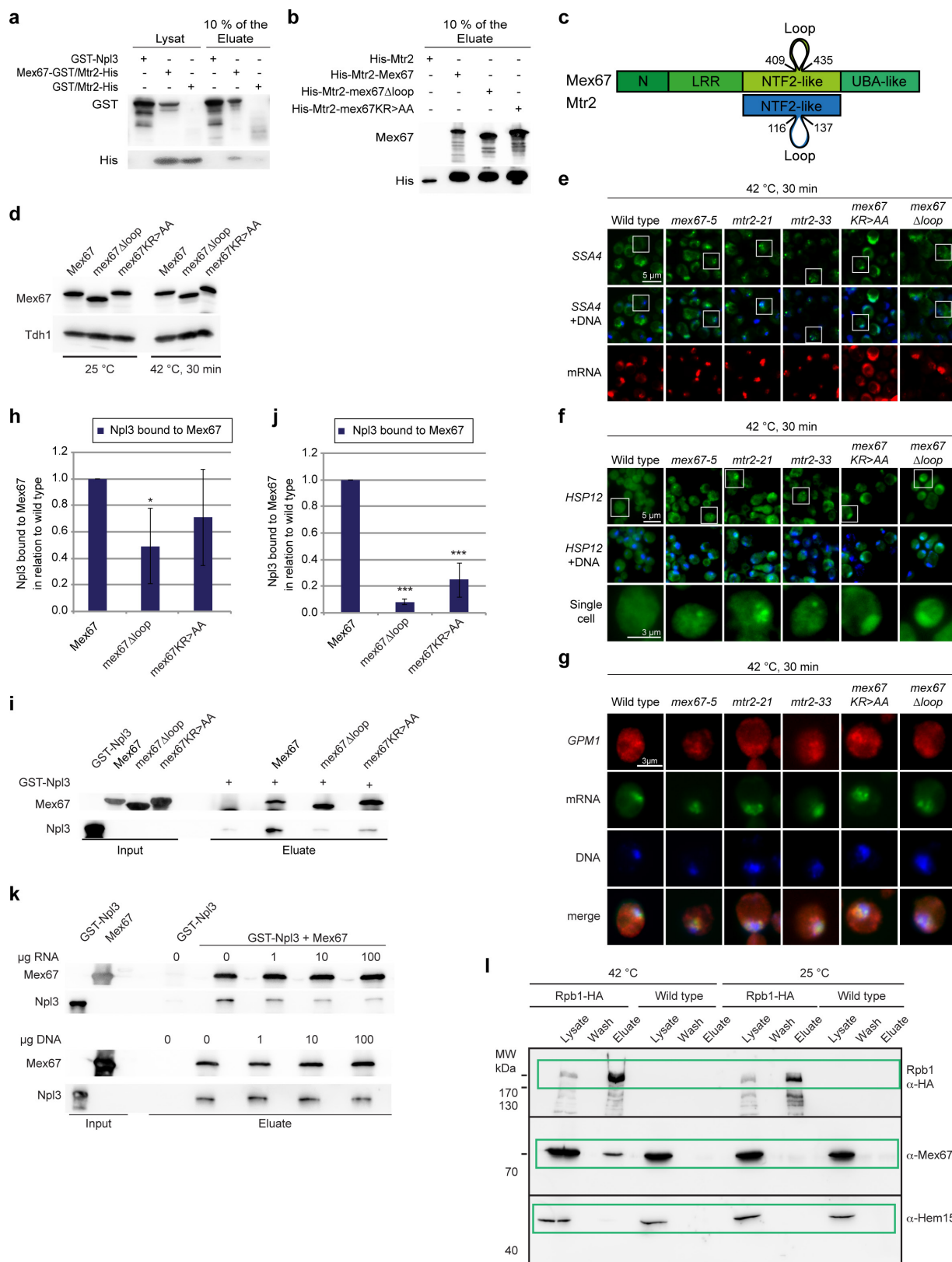
Extended Data Figure 2 | Genome-wide analyses of Mex67- and Npl3-associated RNAs. Dissociation of Mex67 and Npl3 from highly expressed transcripts upon a 30-min heat shock at 42 °C. **a**, Genome-wide RNA-seq studies reflect the dissociation of Mex67 and Npl3 from bulk RNA. RIP experiments were performed with both proteins before and after a 30-min incubation at 42 °C. Shown are 20 highly expressed transcripts arranged into functional groups that were bound by Mex67 or Npl3 and released upon heat shock. **b**, The top 50 transcripts from which indicated

proteins dissociate that were obtained in microarray analysis are displayed by functional group. **c**, **d**, Top 50 transcripts obtained in RNA-seq experiments for Npl3 (**c**) and Mex67 (**d**) are displayed by functional group. **e**, Genome-wide analysis of microarray data showing functional groups of mRNAs that co-purify with Npl3 or Mex67 at 25 °C or 42 °C ($\log_2 \geq 0.5$). **f**, Reduced binding of Npl3 and Mex67 to housekeeping mRNAs (transcripts shown in **e**) upon shift from 25 °C to 42 °C.



Extended Data Figure 3 | Analyses of Mex67 and Npl3 binding to mRNAs of Hsf1 controlled genes. Mex67 binds preferentially to mRNAs controlled by Hsf1 at 42 °C. **a**, Detailed view with the bound Hsf1 gene names depicted shown in Fig. 1c. **b**, Heat map showing the change in binding from 25 °C to 42 °C for Npl3 and Mex67 for the genes depicted

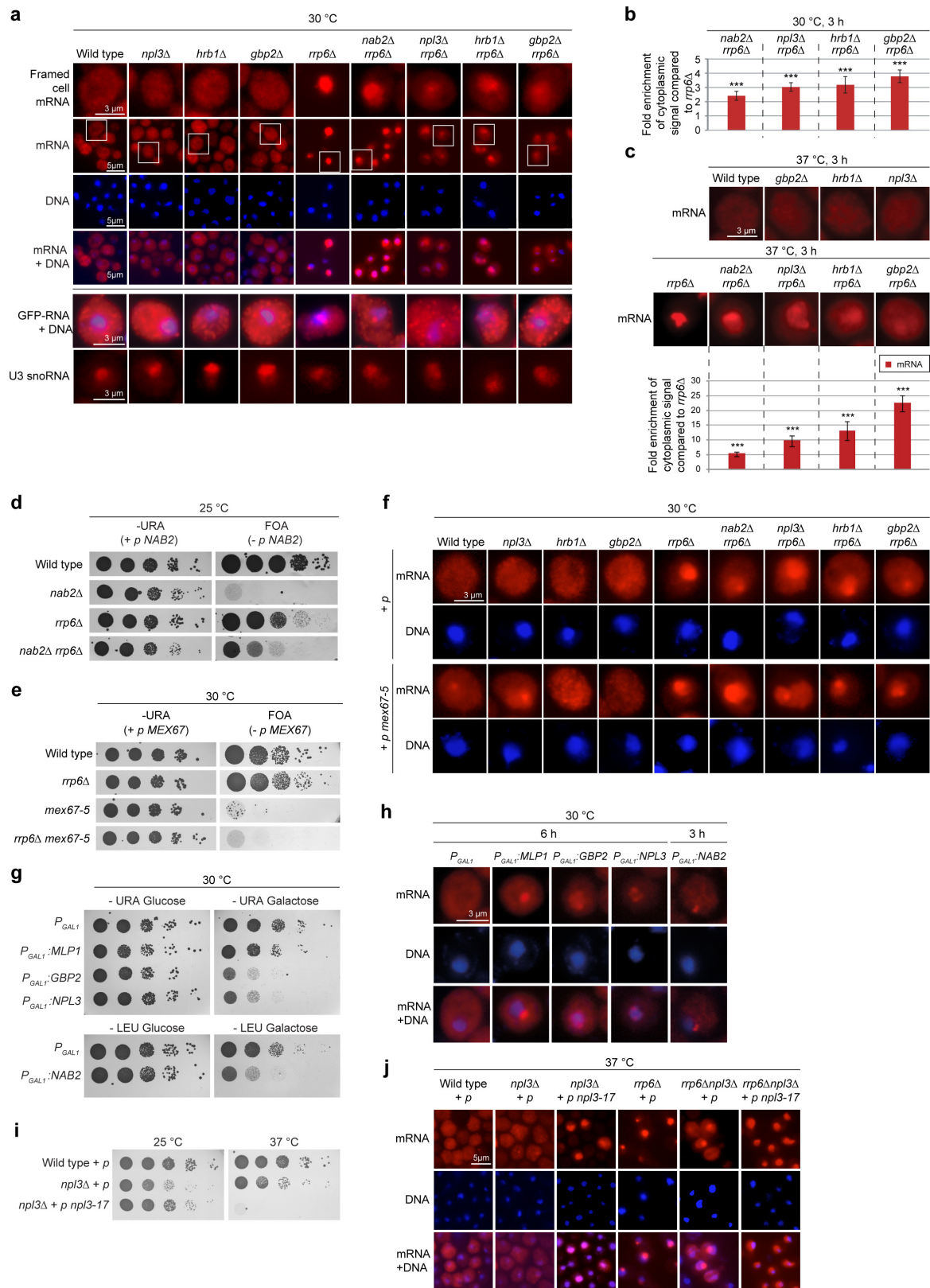
in Fig. 1c. **c**, qRT-PCR experiments for selected transcripts confirm the simultaneous binding of Mex67 and Npl3 to regular transcripts at 25 °C and the selective binding of Mex67 to heat-shock transcripts upon incubation for 30 min at 42 °C.



Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | The loop-domain of Mex67 is important for binding and export of heat-shock mRNAs. **a, b**, Western blot showing purified recombinant His₆-Mtr2 and the co-precipitated recombinant Mex67 with Mex67-specific antibodies (top) and anti-His antibodies (bottom) used for Fig. 2a, b (**a** and **b**, respectively). **c**, Scheme of the heterodimer Mex67-Mtr2 which indicates the loop domain that is mutated in the experiments shown in Fig. 2 (adapted from refs 9, 48). **d**, The loop mutants of Mex67 are not degraded *in vivo*. Western blots are shown from cells cultivated at 25 °C and from cells that were shifted to 42 °C for 30 min. **e, f**, FISH experiments reveal defects in the export of the heat-shock mRNAs *SSA4* (**e**) or *HSP12* (**f**) (green) in mutants of *MEX67* with defects in the loop-domain upon 30-min heat shock at 42 °C. The framed

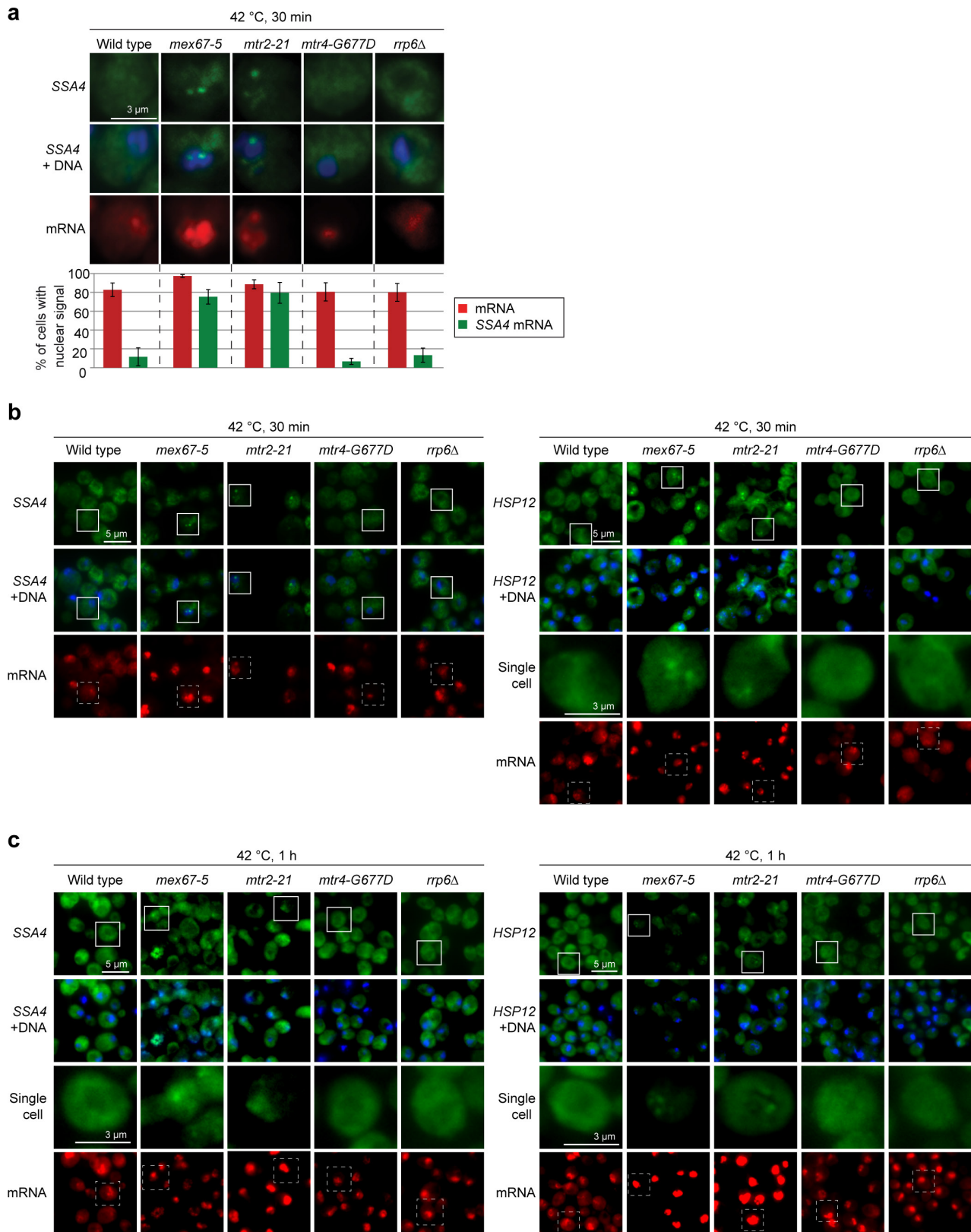
areas in **e** indicate the enlarged cells shown in Fig. 2c. **g**, FISH experiments as performed in **e** and **f**, with probes against the single housekeeping mRNA *GPM1*, show accumulation in all strains. **h**, Quantification of four different experiments shown in Fig. 2d. **i**, *In vitro* interaction study of Npl3 and the indicated Mex67 proteins. Recombinant Mex67 or mutant mex67 were incubated with recombinant GST-Npl3. Subsequent co-immunoprecipitations of Npl3 with Mex67 are shown. **j**, Quantification of five different experiments shown in **i**. **k**, Increasing amounts of RNA, but not DNA, substantially reduce the interaction of Npl3 and Mex67. **l**, Full version of the western blot shown in Fig. 3b. The blot was cut into three pieces (black frame) and probed with the indicated antibodies. The green frame indicates the parts that are shown in Fig. 3b.



Extended Data Figure 5 | See next page for caption.

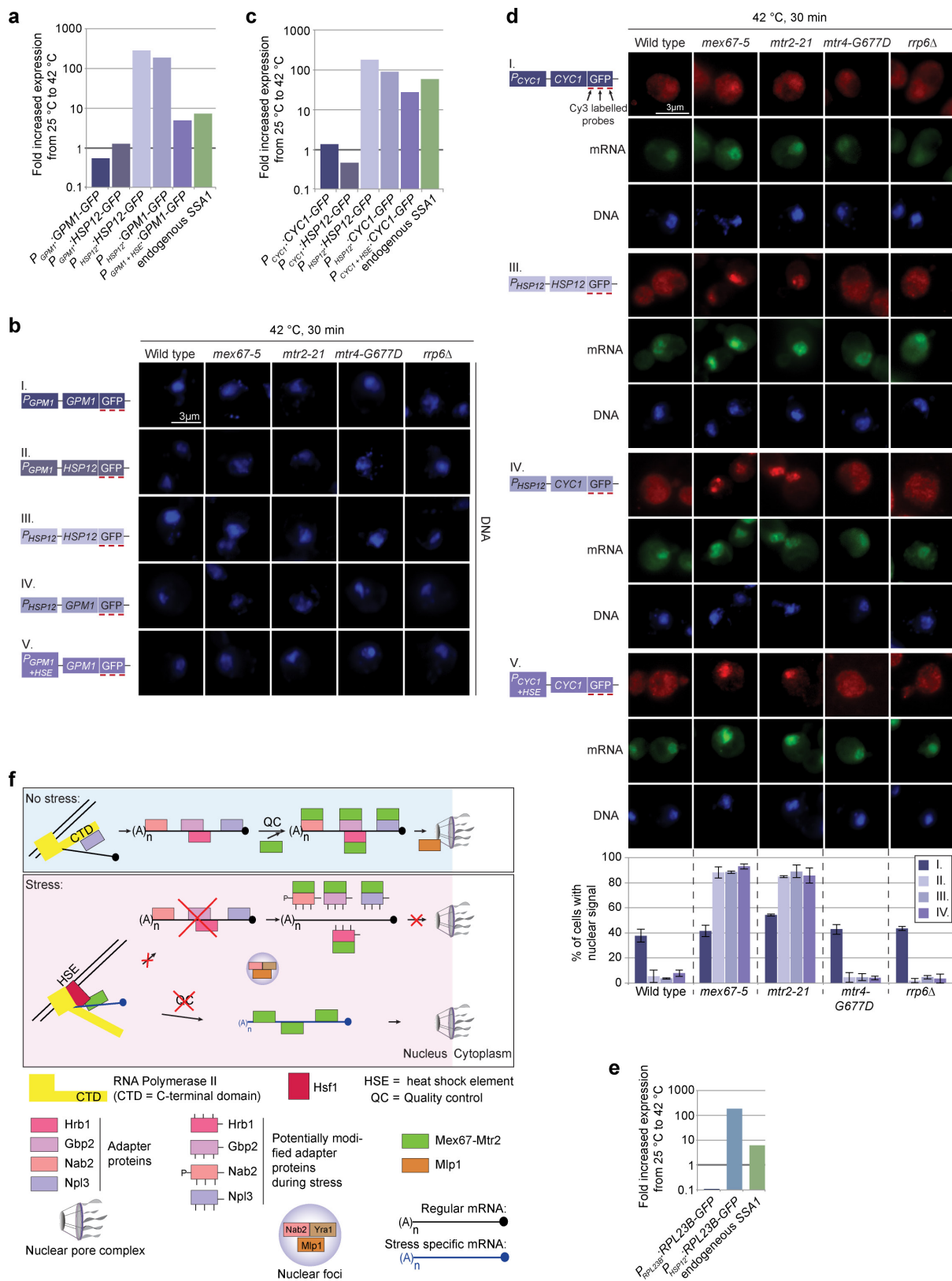
Extended Data Figure 5 | Mex67 adaptor proteins have no mRNA export defects when deleted, but induce defects when overexpressed or mutated. **a, c,** Deletion of the adaptor proteins results in no mRNA export defects and the nuclear accumulation seen in *rrp6* Δ cells alone is reduced when adaptor proteins are also deleted. Cells were grown at 30 °C (**a**) or at 25 °C and shifted to 37 °C for 3 h (**c**) before FISH experiments with an oligo d(T)₅₀ probe (red, top), with a Cy3-labelled probe against an *ADH1* promoter driven *GFP* RNA and the U3 snoRNA (red, bottom) were carried out (**a**). DNA was stained with Hoechst (blue). Framed areas reflect the enlarged cells. **b,** Quantification of FISH experiments shown in Fig. 4a. **d,** The deletion of *NAB2* is lethal, but the simultaneous deletion of *RRP6* suppresses lethality and allows growth at 25 °C. The indicated strains are shown in tenfold serial dilution on plates that retain a covering plasmid or an empty vector (–URA) and on plates that select for the loss of the

covering wild-type plasmid (FOA). **e,** The combination of *rrp6* Δ *mex67-5* is synthetically lethal. Growth of the indicated strains with (–URA) or without (FOA) a covering plasmid is shown. **f,** Overexpression of *mex67-5* in the indicated *rrp6* Δ adaptor protein Δ double mutants decreases the leakage of poly(A)⁺RNA to the cytoplasm as shown by FISH experiments. **g,** Overexpression of quality control factors from the strong *GAL1* promoter is toxic to cells. The indicated strains are shown in tenfold serial dilution on the indicated plates. **h,** Overexpression of quality control factors lead to mRNA export defects and identifies them as mRNA retention factors. FISH with an oligo d(T)₅₀ probe is shown in the indicated strains. **i,** Mutation in *NPL3* is dominant and toxic to cells, as seen in serial drop tests. **j,** Mutation in *NPL3* is dominant and causes mRNA export defects as shown in FISH experiments with an oligo d(T)₅₀ probe.



Extended Data Figure 6 | Heat-shock transcripts are not retained in the nuclei of cells defective in nuclear quality control factors. a, FISH experiments with a DIG-labelled SSA4 probe (green) and an oligo d(T)₅₀ probe (red) are shown in the indicated strains. Nuclear accumulation of

mRNA was quantified (bottom). **b, c,** FISH experiments with DIG-labelled SSA4 and HSP12 probes (green) are shown in the indicated strains that had been shifted to 42 °C for 30 min (**b**) or 1 h (**c**). The frames in **b** indicate the enlarged cells that are also shown in **a**.



Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Transcripts expressed from heat-shock promoters are not quality-controlled. **a**, Representative expression analysis upon heat stress via qRT-PCR. Expression of *GPM1* and *HSP12*, controlled by the indicated promoters, is shown upon 30-min heat stress at 42 °C. The endogenous *SSA1* RNA served as a positive control. **b**, DNA staining of the cells shown in Fig. 4b. **c**, Experiment as shown in **a** but with *CYC1* and *HSP12* transcripts. **d**, Promoter-dependent accumulation of transcripts in quality-control-factor mutants. All strains carrying plasmids with the indicated gene constructs were analysed in FISH experiments with oligonucleotides targeting *GFP* mRNA (red) and poly(A)⁺ RNA (green). The DNA is stained with Hoechst (blue). Examples of typical cells (top) and quantification of the nuclear accumulation (bottom) is shown. **e**, The intron containing *RPL23B* RNA is expressed upon heat stress, when driven from the *HSP12* promoter. qRT-PCR was carried out using lysates of log-phase cells that were subjected to 42 °C for 30 min. **f**, Schematic model of the retention of regular mRNAs under stress and the selective

export of heat-shock mRNAs to the cytoplasm. Under normal conditions, adaptor proteins control the maturation of the transcripts and prevent the early association of Mex67 with Mtr2, avoiding premature nuclear export. Upon maturation, the adaptor proteins recruit Mex67–Mtr2, which allows subsequent nuclear export. Proper Mex67–Mtr2 coverage of the adaptor proteins is controlled at the NPC by the gatekeeper Mlp1 (top). During stress (bottom), regular mRNAs are retained in the nucleus by dissociation of Mex67–Mtr2 through its adaptor proteins. By contrast, heat-shock mRNAs circumvent an adaptor-mediated quality control by instant loading of Mex67–Mtr2 through Hsf1, resulting in an immediate nuclear export. Proper RNP formation is not controlled by Mlp1 at the NPC because Mlp1 is detached and accumulates in nuclear foci. This mechanism allows a quick switch between the controlled export of correct mRNAs and the immediate and uncontrolled export of stress-specific transcripts.

Extended Data Table 1 | Yeast strains used in this study

Number	Genotype	Source
HKY35	<i>MATa ura3-52 leu2Δ1 trp1Δ63</i>	28
HKY36	<i>MATa ura3-52 leu2Δ1 his3-200</i>	28
HKY56	<i>MATa ura3-52 leu2Δ1 rpl1::HIS3</i>	this study
HKY124	<i>MATa ura3-52 leu2Δ1 his3Δ200 rat7-1</i>	29
HKY157	<i>MATa ura3 leu2 his3 ade npl3::HIS3</i> + <i>p CEN URA3 NPL3-myc</i>	30
HKY168	<i>MATa ura3 leu2 trp1 his3 lys2 ade2 ade8 gbp2::HIS</i>	17
HKY194	<i>MATa ura3 leu2 trp1 his3 lys2 ade3 hrb1::HIS3</i>	30
HKY298	<i>MATa ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 hrb1::kanMX4</i>	Euroscarf
HKY307	<i>MATa ura3 leu2 his3 ade2 trp1 yra1::HIS3</i> + <i>p CEN URA3 YRA1</i>	31
HKY369	<i>MATa ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0 gbp2::kanMX4</i>	Euroscarf
HKY380	<i>MATa ura3Δ0 leu2Δ0 his3Δ1 met15Δ0 npl3::kanMX4</i>	Euroscarf
HKY381	<i>MATa ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0</i>	Euroscarf
HKY428	<i>MATa ura3-52 leu2Δ1 his3Δ200 mtr4-G677D</i>	17
HKY642	<i>MATa ura3 leu2 trp1 his3 ade2 mex67::HIS3</i> + <i>p CEN URA3 MEX67</i>	32
HKY643	<i>MATa ura3 leu2 trp1 his3 ade2 mex67::HIS3</i> + <i>p CEN URA3 MEX67</i>	32
HKY644	<i>MATa ura3 leu2 trp1 his3 ade2 mex67::HIS3</i> + <i>p CEN LEU2 mex67-5</i>	32
HKY682	<i>MATa ura3Δ0 leu2Δ0 his3Δ1 npl3::kanMX4</i>	17
HKY719	<i>MATa his3Δ1 leu2Δ0 ura3Δ0 npl3::kanMX4 gbp2::kanMX4 hrb1::kanMX4</i> + <i>p CEN URA3 NPL3</i>	this study
HKY820	<i>MATa ura3-52 leu2Δ1 his3Δ200 GBP2-3xmyc::HIS3</i>	17
HKY821	<i>MATa ura3-52 leu2Δ1 his3Δ200 HRB1-3xmyc::HIS3</i>	17
HKY891	<i>MATa ura3 leu2 his3 lys2 mtr2::kanMX4</i> + <i>p CEN TRP1 mtr2-21 + p CEN URA3 MTR2</i>	33
HKY892	<i>MATa ura3 leu2 his3 lys2 mtr2::kanMX4</i> + <i>p CEN TRP1 mtr2-33 + p CEN URA3 MTR2</i>	33
HKY907	<i>MATa ura3 leu2 his3 mtr2::kanMX4 trp1::kanMX4</i> + <i>p CEN TRP1 mtr2-21</i>	34
HKY931	<i>MATa ura3Δ0 leu2Δ0 his3Δ1 trp1::kanMX4</i>	35
HKY1028	<i>MATa leu2Δ0 his3Δ1 lys2Δ0 ura3Δ0 rrp6::kanMX4</i>	Euroscarf
HKY1115	<i>MATa ura3 leu2Δ nab2::HIS3 + p CEN URA3 nab2ΔN</i>	36
HKY1203	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lys2Δ0 hrb1::HIS3 rrp6::kanMX4</i>	17
HKY1204	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ade2 ade8 trp1 gbp2::HIS3 rrp6::kanMX4</i> + <i>p CEN URA3 GBP2-GFP</i>	17
HKY1263	<i>MATa leu2Δ0 his3Δ1 ura3Δ0 met15Δ0 ssa4::kanMX4</i>	Euroscarf
HKY1266	<i>MATa ura3Δ0 leu2Δ0 his3Δ met15Δ0 MEX67-GFP::HIS3MX6</i>	Invitrogen
HKY1310	<i>MATa leu2Δ0 his3Δ1 lys2Δ0 ura3Δ0 rrp6::kanMX4 npl3::kanMX4</i>	this study
HKY1345	<i>MATa ura3Δ0 leu2Δ0 his3Δ met15Δ0 ADH5-GFP::HIS3MX6</i>	Invitrogen
HKY1409	<i>MATa leu2Δ0 his3Δ1 ura3Δ0 met15Δ0 mlp1::kanMX4</i>	Euroscarf
HKY1457	<i>MATa ura3Δ0 leu2Δ0 his3Δ met15Δ0 HPR1-GFP::HIS3MX6</i>	Invitrogen
HKY1458	<i>MATa ura3Δ0 leu2Δ0 his3Δ met15Δ0 YRA1-GFP::HIS3MX6</i>	Invitrogen
HKY1475	<i>MATa ura3Δ0 leu2Δ0 his3Δ met15Δ0 NAB2-GFP::HIS3MX6</i>	Invitrogen
HKY1506	<i>MATa ura3 leu2 his3Δ1 nab2::HIS3 rrp6::kanMX4</i> + <i>p CEN URA3 nab2ΔN</i>	this study
HKY1568	<i>MATa ura3Δ0 leu2Δ0 his3Δ met15Δ0 CYC1-GFP::HIS3MX6</i>	Invitrogen
HKY1569	<i>MATa ura3 leu2 trp1 his3 ade2 mex67::HIS3</i> + <i>p CEN LEU2 mex67Δ409-435aa</i>	this study
HKY1570	<i>MATa ura3 leu2 trp1 his3 ade2 mex67::HIS3</i> + <i>p CEN LEU2 mex67KR>AA</i>	this study
HKY1571	<i>MATa ura3Δ0 leu2Δ0 rrp6::kanMX4 CYC1-GFP::HIS3MX6</i>	this study
HKY1572	<i>MATa ura3Δ0 leu2Δ0 mtr4-G677D CYC1-GFP::HIS3MX6</i>	this study
HKY1580	<i>MATa ura3Δ0 leu2Δ0 his3Δ met15Δ0 HSF1-GFP::HIS3MX6</i>	Invitrogen
HKY1640	<i>MATa mex67::HIS3 CYC1-GFP::HIS3MX6 ura3Δ0 leu2Δ0</i> + <i>p CEN LEU2 mex67-5 + p CEN URA3 P_{CYC1}:CYC1-GFP</i>	this study
HKY1641	<i>MATa mtr2::kanMX4 CYC1-GFP::HIS3MX6 trp1::kanMX4 ura3Δ0 leu2Δ0</i> + <i>p CEN TRP1 mtr2-21 + p CEN URA3 P_{CYC1}:CYC1-GFP</i>	this study
HKY1689	<i>MATa rrp6::kanMX4 mex67::HIS3</i> + <i>p CEN LEU2 mex67-5 + p CEN URA MEX67</i>	this study

Extended Data Table 2 | Oligonucleotides used in this study

Number	Sequence	Name
HK818	5'-CCACTATAACACCCGATGAGG-3'	<i>IDP3</i> forward
HK819	5'-GGCAAATAAGGTTACAGTTTAGC-3'	<i>IDP3</i> reverse
HK1002	5'-TGCTAAGGCTGTCGGTAAGG-3'	<i>TDH1</i> forward
HK1003	5'-TCAGAGGAGACAACGGCATC-3'	<i>TDH1</i> reverse
HK1066	5'-GGTTGAACGGTTCTTGTATGGC-3'	<i>ADH1</i> forward
HK1067	5'-CCAAAGAACCTAGACCACCAAGC-3'	<i>ADH1</i> reverse
HK1451	5'-AGGCTCGTAGCGGTTCTGAC-3'	<i>25S rRNA</i> forward
HK1452	5'-CGAGCTTCTGCTATCCTGAGGG-3'	<i>25S rRNA</i> reverse
HK1473	5'-CACCTACGCTGACAACCAACC-3'	<i>SSA4</i> forward
HK1474	5'-CATCCTCTTCACCCACCTTCTCC-3'	<i>SSA4</i> reverse
HK1475	5'-CAAGGATAACGCTGAAGGTCAAGG-3'	<i>HSP12</i> forward
HK1476	5'-CTTCTTGGTTGGGTCTTCTTACC-3'	<i>HSP12</i> reverse
HK1477	5'-GTTGAAAGTCGTGGTTCTGTTG-3'	<i>HSP26</i> forward
HK1478	5'-CTGCTCTCCTTGACCTTGACCTTG-3'	<i>HSP26</i> reverse
HK1494	5'-AAAGCACCGTTTCCCGTCC-3'	<i>5S rRNA</i> forward
HK1495	5'-CACTACACTACTCGGTCAGGCT-3'	<i>5S rRNA</i> reverse
HK1509	5'-CGGGCTCTGGAAGAATGTGTTG-3'	<i>GRE3</i> forward
HK1510	5'-GAGGACCGAAGGAGGAGTAAGC-3'	<i>GRE3</i> reverse
HK1511	5'-GTCTTCTCCGCTCAAACCTCC-3'	<i>SSA1</i> forward
HK1512	5'-GAACAGCAGCACCGTAAGCAAC-3'	<i>SSA1</i> reverse
HK1513	5'-GCCCAATGCAAGAAACCAAC-3'	<i>EDC2</i> forward
HK1514	5'-CCATCGTTGCTTATCGTCCTC-3'	<i>EDC2</i> reverse
HK1598	5'-GGCCCCAGGTAAGAAAGTCG-3'	<i>RPL8A</i> forward
HK1599	5'-GAAGGTTTCGGCAGCGGTG-3'	<i>RPL8A</i> reverse
HK1921	5'-AAGATGGCGTGAAGAAGGCA-3'	<i>HEM15</i> forward
HK1922	5'-CCGCAACCTGTCAGAGACAA-3'	<i>HEM15</i> reverse
HK1955	5'-GGTGAGCCAGGTATCGGTAAGAC-3'	<i>HSP104</i> forward
HK1956	5'-CCGATGACCTTCAATTGGCCTC-3'	<i>HSP104</i> reverse
HK2040	5'-Cy3-CCATTAACATCACCATCTAATTCAACAAGAATTGGGACAACCTCCAGTGAA-Cy3-3'	<i>GFP</i> reverse
HK2041	5'-Cy3-CTTGACTTCAGCACGTGTCTTGTAGTTCCTGTCATCTTTGAAAAATATAG-Cy3-3'	<i>GFP</i> reverse
HK2055	5'-Cy3-TAAAAGGACAGGGCCATCGCCAATTGGAGTATTTGTTGATAATGGTCTGCT-Cy3-3'	<i>GFP</i> reverse
HK2098	5'-CCGGGCAGAACATTCTAGAAAGC-3'	<i>HSE</i> forward
HK2099	5'-CCGGGCTTTCTAGAAATGTTCTGC-3'	<i>HSE</i> reverse
HK2128	5'-Cy3-TTCTTTAAAATCAATACCTTTTAACTCGATTCTATTAACAAGGGTATCAC-Cy3-3'	<i>GFP</i> reverse
HK2129	5'-Cy3-TCCGGGTATCTTGAAGAGCTGAACACCATAAGTGAAAGTAGTGACAAG-Cy3-3'	<i>GFP</i> reverse
HK2134	5'-ATGCCCGAAGGTTATGTACAGG-3'	<i>GFP</i> forward
HK2135	5'-CATTCTTTTGTGTTGCTGCCATG-3'	<i>GFP</i> reverse
HK2138	5'-GGTGCCTTCAACATTACTAGACTGC-3'	<i>NTR Chr.V</i> forward
HK2139	5'-GATATGTATATGCGCAAGAAGGTGC-3'	<i>NTR Chr.V</i> reverse
HK2150	5'-CGCAGGTAGAAAAGGATTTCGG-3'	<i>HSP12</i> forward
HK2151	5'-CCTTTTTCGGCAGAGTCG-3'	<i>HSP12</i> reverse
HK2154	5'-CCAGAACAATCCGTACACAAGG-3'	<i>HEM15</i> forward
HK2155	5'-GCAATTGTCTTCTGATACTTAGCAC-3'	<i>HEM15</i> reverse
HK2170	5'-Cy3-CTCTTTTCGTTGGGATCTTTCGAAAGGGCAGATTGTGTGGACAGGTAATG-Cy3-3'	<i>GFP</i> reverse
HK2171	5'-Cy3-CCTGTACATAACCTTCGGGCATGGCACTCTTGAAAAAGTCATGCCGTTTC-Cy3-3'	<i>GFP</i> reverse
HK2226	5'-GGTTGGCAACAGCAGCGGCACAGCAGCGGCAGCTTCTGGGTCCAAGTAG-Cy3-3'	<i>GPM1</i> reverse
HK2227	5'-CCAATGGAATACCAGTTGGGATGTTCAACTTAGCAATGTCAGCATCAGAG-Cy3-3'	<i>GPM1</i> reverse
HK2288	5'-CY3-ATTCAGTGGCTCTTTTGAAGAGTCAAAGAGTGACGATTCTATAGAAATGA-3'	<i>U3 snoRNA</i>
HK2354	5'-GGAGAACAAGGGCTCCAGATTG-3'	<i>HSP30</i> forward
HK2355	5'-CCAGTAGTACTAGCGGCTAACTCG-3'	<i>HSP30</i> reverse
HK2356	5'-GGACGGATCGGCATTGCC-3'	<i>AHA1</i> forward
HK2357	5'-CACCTGAGATTCGGGCACC-3'	<i>AHA1</i> reverse
HK2360	5'-GGTGTCCACCATATGAAGGCAC-3'	<i>HSP42</i> forward
HK2361	5'-GGTAAGTCCATTCTCGTTTCAGGC-3'	<i>HSP42</i> reverse
HK2370	5'-GGTCAAAGACGCTGAATCTGC-3'	<i>SGT2</i> forward
HK2371	5'-CGACATCAGCATCCCTTGACTG-3'	<i>SGT2</i> reverse
HK2380	5'-CGATAATGATGACAGCAGCAGC-3'	<i>CTH1</i> forward
HK2381	5'-GGTTAATCTCAACTTCGAGTGACC-3'	<i>CTH1</i> reverse
HK2429	5'-GGACAAGCTTTTATCCGTTGACG-3'	<i>ERB1</i> forward
HK2439	5'-CCATCTGCATATTTGAGTAAATATTCGG-3'	<i>ERB1</i> reverse
HK2433	5'-GGAGCAGGAAGAAATGATGGC-3'	<i>NOP2</i> forward
HK2434	5'-GGTTCTGATGGTTATTGGTCTTCG-3'	<i>NOP2</i> reverse
HK2446	5'-CAGAGATGTACGACTATCGC-3'	<i>PIB1</i> forward
HK2447	5'-CATGCCGTACCGGATTG-3'	<i>PIB1</i> reverse

Extended Data Table 3 | Plasmids used in this study

Number	Features	Source
pHK12	<i>CEN URA3 P_{ADH} NLS-NES-GFP-GFP</i>	37
pHK20	<i>CEN LEU2 MEX67-GFP</i>	32
pHK79	<i>CEN URA3 P_{GALI}</i>	this study
pHK80	<i>CEN LEU P_{GALI}</i>	this study
pHK87	<i>CEN LEU2</i>	38
pHK88	<i>CEN URA3</i>	38
pHK103	<i>2μ LEU2</i>	39
pHK104	<i>2μ URA3</i>	39
pHK154	<i>CEN LEU2 npl3-17</i>	40
pHK144	<i>2μ URA P_{GALI} GFP-NPL3</i>	41
pHK291	<i>CEN LEU2 mex67-5-ProtA</i>	42
pHK358	<i>CEN LEU2 P_{GALI} -NAB2-GFP</i>	43
pHK367	<i>CEN URA3 GBP2-GFP</i>	24
pHK422	<i>2μ URA P_{GALI} GBP2-GFP</i>	24
pHK418	<i>CEN LEU2 GFP-NPL3</i>	44
pHK439	<i>GST</i>	45
pHK442	<i>GST-MEX67</i>	45
pHK477	<i>CEN TRP yra1-1</i>	31
pHK531	<i>CEN LEU2 RPB1-HA</i>	46
pHK765	<i>CEN URA3 GFP-NPL3</i>	17
pHK778	<i>CEN LEU2 9xmyc-NPL3</i>	34
pHK779	<i>CEN URA3 9xmyc-NPL3</i>	17
pHK892	<i>CEN URA3 P_{ADH} PRP17-GFP</i>	17
pHK927	<i>CEN URA3 P_{ADH} HRB1-GFP</i>	17
pHK1221	<i>2μ URA P_{GALI} MLP1</i>	47
pHK1276	<i>GST-NPL3</i>	35
pHK1279	<i>P_{Trc}:HIS:MTR2</i>	this study
pHK1372	<i>P_{Trc}:HIS:TEV:MTR2:MEX67</i>	9
pHK1373	<i>P_{Trc}:HIS:TEV:MTR2:mex67loopKR>AA</i>	9
pHK1374	<i>P_{Trc}:HIS:TEV:MTR2:mex67-409-435aaK343E</i>	9
pHK1376	<i>CEN LEU2 mex67Δ409-435aa</i>	9
pHK1377	<i>CEN LEU2 mex67loopKR>AA (K415A, K416A, K419A, K424A, R426A, R427A)</i>	9
pHK1443	<i>CEN URA3 P_{CYC1}:CYC1-GFP</i>	this study
pHK1444	<i>CEN URA3 P_{HSP12}:CYC1-GFP</i>	this study
pHK1445	<i>CEN URA3 P_{HSP12}:HSP12-GFP</i>	this study
pHK1464	<i>CEN URA3 P_{CYC1+HSE}:CYC1-GFP</i>	this study
pHK1470	<i>CEN URA3 P_{GPM1}:GPM1-GFP</i>	this study
pHK1472	<i>CEN URA3 P_{HSP12}:GPM1-GFP</i>	this study
pHK1517	<i>CEN URA3 P_{CYC1}:HSP12-GFP</i>	this study
pHK1518	<i>CEN URA3 P_{GPM1}:HSP12-GFP</i>	this study
pHK1547	<i>CEN URA3 P_{HSP12}:RPL23B-GFP</i>	this study
pHK1548	<i>CEN URA3 P_{RPL23B}:RPL23B-GFP</i>	this study
pHK1553	<i>CEN URA3 P_{GPM1+HSE}:GPM1-GFP</i>	this study

Near-atomic-resolution cryo-EM analysis of the *Salmonella* T3S injectisome basal body

L. J. Worrall^{1*}, C. Hong^{2*}, M. Vuckovic¹, W. Deng³, J. R. C. Bergeron^{1†}, D. D. Majewski¹, R. K. Huang², T. Spreter^{1†}, B. B. Finlay³, Z. Yu² & N. C. J. Strynadka¹

The type III secretion (T3S) injectisome is a specialized protein nanomachine that is critical for the pathogenicity of many Gram-negative bacteria, including purveyors of plague, typhoid fever, whooping cough, sexually transmitted infections and major nosocomial infections. This syringe-shaped 3.5-MDa macromolecular assembly spans both bacterial membranes and that of the infected host cell. The internal channel formed by the injectisome allows for the direct delivery of partially unfolded virulence effectors into the host cytoplasm¹. The structural foundation of the injectisome is the basal body, a molecular lock-nut structure composed predominantly of three proteins that form highly oligomerized concentric rings spanning the inner and outer membranes^{2–5}. Here we present the structure of the prototypical *Salmonella enterica* serovar Typhimurium pathogenicity island 1 basal body, determined using single-particle cryo-electron microscopy, with the inner-membrane-ring and outer-membrane-ring oligomers defined at 4.3 Å and 3.6 Å resolution, respectively. This work presents the first, to our knowledge, high-resolution structural characterization of the major components of the basal body in the assembled state, including that of the widespread class of outer-membrane portals known as secretins.

Using single-particle cryo-electron microscopy (cryo-EM), we determined the structure of a secretion-incompetent basal body mutant, lacking the N-terminal cytoplasmic domain of PrgH² (PrgH_{130–392}), at a resolution of 6.3 Å (Fig. 1a and Extended Data Fig. 1a–d, g). In our structure, the periplasmic gate is in a closed conformation (Fig. 1c) and, by design, no density is attributable to the inner rod or needle, neither of which was detected by liquid chromatography–tandem mass spectrometry (LC–MS/MS) (Extended Data Fig. 2a). The inner-membrane rings (*Salmonella* pathogenicity island 1 (SPI-1) proteins PrgH and PrgK) have the best local resolution, whereas the export apparatus components they encompass and the outer-membrane pore (InvG, a member of the secretin family that is also found in the type II secretion system (T2SS), type IV pilus system and the phage-assembly system⁶) are of lower resolution (Extended Data Fig. 1i). Imposing 24-fold symmetry^{5,7} improved the global resolution to 4.3 Å (Extended Data Fig. 1e, f, h, j), resolving side chains for the inner-membrane rings (Fig. 1b and Extended Data Fig. 1k) and allowing for models of PrgH_{171–364} and PrgK_{20–203} to be built. To probe the structure of the InvG secretin further, we purified the intact protein alone⁸, with cryo-EM analysis giving an unsymmetrized map of 4.4 Å resolution (Extended Data Fig. 3a–e) and of 3.6 Å with clear 15-fold symmetry imposed (this allowed the structurally uncharacterized N3, secretion and S domains (InvG_{172–557}) to be built; Fig. 1b and Extended Data Fig. 3f–i). Superimposition with the basal body map confirms highly similar overall features (Fig. 1c), indicating that no major conformational artefacts were introduced during secretin isolation. In the absence of the inner-membrane rings, the terminal N0

and N1 domains of InvG were not resolved; we therefore employed Rosetta electron-microscopy-guided symmetrical modelling² to position the previously determined X-ray structure² in the 6.3 Å basal body map (Fig. 1d).

The inner-membrane proteins PrgH and PrgK form concentric rings (Fig. 1d), consistent with our earlier Rosetta electron-microscopy-guided models^{2–4} but with more closely packed homo- and hetero-oligomerization interfaces (Extended Data Fig. 2b–e). Notably, the uninterrupted sequence of PrgK_{20–203} is observed here, including two previously unresolved regions that we now understand to be central to the assembly of PrgH and PrgK. First, the linker connecting the D1 and D2 domains (Gln76–Pro98) packs between neighbouring PrgK D2 domain helices, with conserved residue Phe89 (the mutation of which abrogates self-association³) inserted in a zipper-like fashion around the assembled ring (Extended Data Fig. 2d). Second, the loop connecting the PrgK D2 domain and the C-terminal transmembrane helix forms extensive interactions at the interface of neighbouring PrgK and PrgH monomers (Extended Data Fig. 2e), a finding supported by earlier cross-linking data⁹. A region of this loop and the C-terminal transmembrane helix are absent in some bacterial variants, such as that of the enteropathogenic *Escherichia coli* (EPEC) protein EscJ (terminates at the PrgK₁₉₅ equivalent). Although it still allows for secretion, a PrgK_{1–200} mutant destabilizes the needle complex³, suggesting that this region contributes to a more robust assembly, one that may be required depending on the particular environmental niche of the pathogen.

Localized within the bacterial outer membrane, secretins are large portals to the extracellular environment essential in four distinct bacterial secretion systems⁶. The structure of InvG_{172–557} is a massive double-layered β-barrel (Fig. 2 and Extended Data Fig. 4). The central secretin domain (residues 302–518) forms an eight-stranded β-sandwich that is splayed apart at the extracellular end (Fig. 2c) and is highly conserved at the packed core (Extended Data Figs 5, 6). The outer β-sheet (strands 1, 3, 8 and 9) forms a 60-stranded anti-parallel β-barrel (with a shear of 60) that constitutes the outer wall; the inner sheet (strands 4–7) forms a kinked anti-parallel barrel (strand 60; shear 0) that both buttresses the outer wall and forms the inner ‘periplasmic gate’ (Fig. 2b, c). At the extracellular face of the outer wall, strands 1 and 2 form a hairpin that, along with a kinked strand 3, constitutes an angled 45-stranded amphipathic β-barrel ‘lip’, decorated on the exterior by a highly conserved amphipathic helical loop that connects strands 8 and 9 (Extended Data Figs 5b, 6). The preceding N3 domain abuts the periplasmic base of the secretin domain. It belongs to the family of small mixed α/β-modular domains that we previously termed ring-building motifs (RBMs)⁴, common to proteins of the injectisome² and other secretion systems¹⁰. These domains share an overall superficial architecture (divided into two groups based on secondary structure connectivity) and, on the basis of prior modelling, a broadly conserved ring-packing arrangement is predicted². The structure of the 15-mer

¹Department of Biochemistry and Molecular Biology and the Center for Blood Research, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada. ²CryoEM Shared Resources, Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147, USA. ³Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. [†]Present addresses: Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA (J.R.C.B.); Zymeworks, Vancouver, British Columbia V6H 3V9, Canada (T.S.).

*These authors contributed equally to this work.

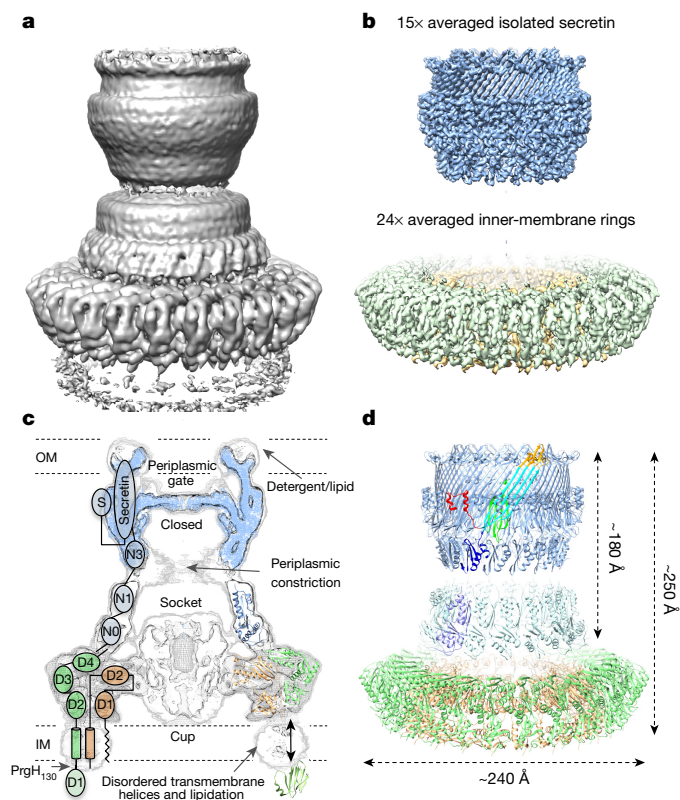


Figure 1 | Cryo-EM structures of the injectisome basal body and isolated secretin. **a**, Reconstruction of the PrgH_{130–392} basal body (C1) at 6.3 Å resolution. **b**, Reconstructions of the 15-fold averaged isolated secretin (top: blue; 3.6 Å resolution) and the 24-fold averaged inner-membrane rings of the basal body map (bottom: PrgH, green; PrgK, orange; 4.3 Å resolution). **c**, Central slice view of basal body reconstruction (dark-grey contoured as in **a** and light-grey contoured at lower level to highlight less-ordered features) and isolated secretin (blue). The domain annotation of PrgH, PrgK and InvG is overlaid on the left and the structures of the monomeric domains previously solved on the right. The PrgH cytoplasmic D1 domain (green, bottom left) is not present in the PrgH_{130–392} mutant used in this study and its precise location with respect to the basal body is unclear. The transmembrane helices of PrgH (N-terminal) and PrgK (C-terminal) and the PrgK N-terminal lipidation are present but diffusely ordered. **d**, Refined structures for InvG_{172–557} (blue), PrgH_{171–364} (green), PrgK_{20–203} (orange) and Rosetta-modelled InvG_{34–171} (pale blue). One monomer encompassing InvG_{34–557} is coloured according to structural domains: medium blue, N0–N1 domains; cobalt blue, N3 domain; cyan, outer β -sheet; green, inner β -sheet; orange, secretin domain lip; red, S domain (note the displaced interaction with the β -sheet of the $i + 1$ and $i + 2$ protomers).

N3 domain here (as well as the 24-mer RBMs of PrgH and PrgK) validates this oligomerization mechanism and highlights the substantial network of interactions that these domains make to the overall assembly and stability of the injectisome. A β -hairpin specific to T3S secretin N3 domains (residues 193–206) packs against the underside of the β -sheets of the periplasmic gate (Fig. 2c). C-terminal to the secretin domain, the S domain (residues 526–end) forms a helix-turn-helix motif that extends laterally across neighbouring protomers at the outer midsection of the secretin pore (Figs 1d, 2).

The InvG secretin structure unambiguously confirms the pentameric stoichiometry previously suggested in 10 Å resolution cryo-EM maps of the *Salmonella* SPI-1 injectisome⁸. Variation in the symmetry of other secretins, ranging from 12- to 14-fold, has been proposed, but is unexplained by our structure-guided sequence analysis, with the double-layered barrel well conserved and most variation involving loop regions that have limited involvement in oligomerization (Extended Data Fig. 6). Similarly, comparison of previous electron microscopy

reconstructions does not reveal a relationship between proposed stoichiometry and secretin dimensions, with the InvG 15-mer here fitting closely to lower-resolution maps from other species (Extended Data Fig. 7d, e).

The InvG structure provides a molecular framework that allows us to understand the features that contribute to function of the secretin family, including assembly, localization, membrane association and gating. It also provides context to the wealth of published biochemical, biophysical and mutational *in vitro* and *in vivo* experimental data.

The observed InvG domains contribute extensively to oligomerization, a property that is mediated by a significant hydrophobic interface (a solvation free-energy gain (Δ^iG) of $-45 \text{ kcal mol}^{-1}$ (ref. 11) and a surface area of approximately $4,700 \text{ Å}^2$ per monomer), in keeping with the highly thermostable and denaturation-resistant signatures of these pores in the assembled state^{12,13}. In addition to the extensive β -strand hydrogen-bond network of the central secretin domain (Extended Data Fig. 7a), the smaller N3 and S domains also make notable contributions to oligomerization, and their disposition is suggestive of a concerted role in assembly (Extended Data Fig. 7b, c). The N3 RBM forms not only multi-valent homo-oligomeric ring contacts (with an interface area of $\sim 770 \text{ Å}^2$ per monomer), but also interfaces with the β -sandwich of the adjacent secretin domain ($\sim 550 \text{ Å}^2$ per monomer), which explains why it is essential in oligomerization (as opposed to the more N-terminal N0, N1 and N2 domains, which have been shown to be dispensable for secretin assembly in PulD)¹⁴. As validation, PulD oligomerization-deficient mutants¹⁵ map to both the observed N3 self-association and N3-secretin domain interfaces (Extended Data Fig. 5d), while Leu293Arg mutation of InvG at the core of the N3 ring interface reduces secretion and stability (Extended Data Fig. 5f). Notably, the N3 self-association interface is largely hydrophobic ($\Delta^iG = -16.4 \text{ kcal mol}^{-1}$), typical of a very stable complex. By comparison, the analogous RBM interface formed by the more peripheral InvG N1 domain ($\Delta^iG = -3.6 \text{ kcal mol}^{-1}$) has a more polar, loosely associated interface, suggestive of plasticity in basal body assembly. The differing numbers of N-terminally disposed domains amongst the secretin class (T2SS PulD has an additional module (N2) compared to type III secretion system (T3SS) variants, for example), suggest the need for a customized periplasmic span that couples each particular secretin to the varying inner-membrane components of these systems.

The C-terminal S domain wraps around the secretin midsection, forming interactions with an extended hydrophobic surface on the exterior of the β -barrel that spans over two successive monomers (an interface area of $\sim 1,500 \text{ Å}^2$) in a stabilizing role, akin to that of a molecular staple (Extended Data Fig. 7b, c). PulD multimerization-deficient mutants¹⁵ map to this interface (Extended Data Fig. 5d), supporting a role for this region in assembly. The S domain is also the site of interaction with cognate chaperones known as pilotins: small outer-membrane targeted lipoproteins that are involved in secretin localization, assembly and outer-membrane insertion¹⁶. Pilotins are structurally diverse but appear to interact with their respective secretin in a conserved manner, as illustrated by the structures of the *Shigella flexneri* T3S MxiM¹⁷ and the *Dickeya dadantii* T2S OutS¹⁸ pilotins in complex with secretin S domain peptides. In both, a helical structure of the peptide forms only upon binding to the hydrophobic surface of the pilotin, a common disorder-order transition that, according to our data, also occurs for the complex of InvG and the *Salmonella* pilotin InvH (Extended Data Fig. 8c). Furthermore, the high affinity of the secretin-pilotin interaction (Extended Data Fig. 8d) is consistent with the pilotin-mediated outer-membrane targeting of InvG monomers, potentially with assistance from the bacterial Lol lipoprotein-sorting pathway. The InvG S domain and the S domain peptide of the *S. flexneri* secretin MxiD share a common turn-helix motif (Extended Data Fig. 8a) and superimposition localizes the pilotin to the exterior of the InvG multimer (Extended Data Fig. 8b). Deletion of this motif (residues 542–562) in InvG results in the same secretion-deficient phenotype as removal of the complete S domain (residues 521–562) (Extended Data Fig. 8g). Mutation at

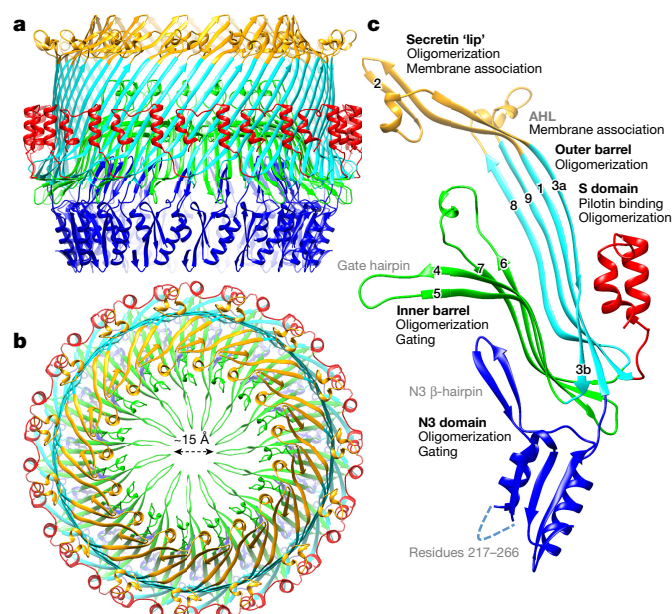


Figure 2 | Structure of the InvG secretin pore. **a–c**, InvG_{172–557} secretin pore structure viewed from the side (**a**), the top (outer-membrane perspective; **b**) and the monomeric (**c**) view. Domains are coloured according to structural features: blue, N3 oligomerization ring domain; cyan, secretin domain outer β -sheet; green, secretin domain inner β -sheet; gold, secretin membrane association, AHL and membrane insertion domain lip; red, C-terminal pilotin-binding S domain. Secretin domain β -strand numbers and functional roles of domains are indicated in **c**.

the modelled pilotin interface results in an approximately 500-fold reduction in binding affinity of the isolated S domain and InvH, with a corresponding diminished secretion phenotype (Extended Data Fig. 8d, e, g). Together, the data support a stabilizing role for the pilotin: it acts as a molecular 'stapler' that mediates the folding of the S domain, which is necessary to allow the observed clamping interactions across two preceding secretin protomers ($i + 1$ and $i + 2$) during assembly (Fig. 1d and Extended Data Fig. 7b). This is consistent with the observation that the T2SS pilotin PulS substantially catalyses the initial stages of PulD folding¹⁹. Previous *in situ* tomography analysis of *S. flexneri*, together with this latter work, suggests that the pilotin may stay associated with secretin in the assembled injectisome²⁰ (Extended Data Fig. 9d). Collectively, our observed contributions of both the ring-forming N3 and staple-like S domains to oligomerization indicate that each has a key role in efficient secretin assembly and stability.

Secretins from several systems^{19,21,22} have been shown to spontaneously insert into the bacterial outer membrane, independent of the general β -barrel assembly machinery (BAM). Biophysical data, including low-resolution electron microscopy of T2S PulD, helped to further define this process as step-wise through the outer-membrane attachment of a 'pre-pore'—that is, a partially ordered oligomer that we can now structurally define as being comprised of a stabilized core of the N3, secretin and S domains described above but with unstructured distal (outer-membrane) secretin domain β -sheets (Extended Data Fig. 9c)—followed by insertion of the folded pore into the membrane^{19,23,24}. The structure of InvG delineates the structural motifs implicated in these distinct membrane association and insertion steps. The only considerable hydrophobic surface is localized at the outer face of the kinked β -barrel lip, with a span and composition consistent with that of a membrane-spanning domain (Fig. 3a–c). A diffuse ring of density surrounds this region, similar to the detergent/lipid belt around transmembrane domains of other proteins imaged by cryo-EM. A notable feature of the transmembrane domain is the dramatic angle of the β -strands, which are kinked by approximately 45° relative to the outer-membrane surface normal (Fig. 3a). This matches well with

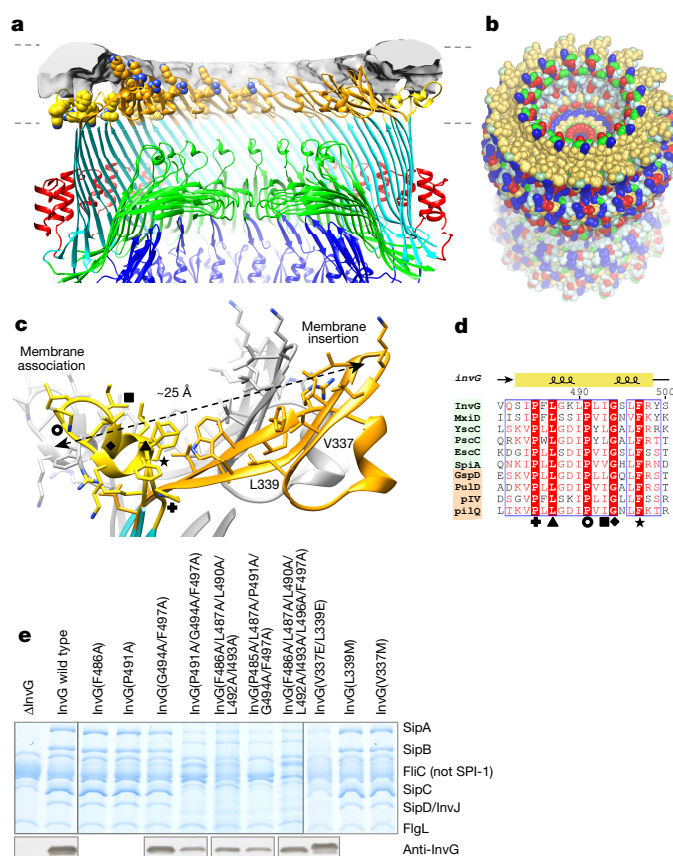


Figure 3 | Structural elements mediating secretin membrane association and insertion. **a**, Slab view of the secretin with AHL (gold) that we propose to mediate the initial monotopic association with the inner leaflet of the outer membrane and the amphipathic kinked β -barrel lip (orange) that we propose allows for subsequent full BAM-independent insertion of the pore across the bilayer. Aromatic and positively charged residues are shown as spheres on monomers to the left. The density we attribute to diffusely ordered detergent and lipid in the C1 secretin map is shown as a grey surface. **b**, Solvent-accessible surface of the secretin (including N0 and N1 domains) is coloured according to residue type: pale orange, hydrophobic residues; cyan, polar; blue, positively charged; red, negatively charged; green, glycine; light green, proline. **c**, Close-up view of the proposed membrane-spanning region of InvG formed by the kinked region of strands 1, 2, 3a (orange; residues 321–358) and the conserved AHL that links strands 8 and 9 (gold; residues 483–498). Conserved residues from **d** are labelled, including PulD multimerization mutants Ile569Ser (InvG Ile493 equivalent; solid square) and Phe573Leu (InvG Phe497 equivalent; solid star). **d**, Sequence alignment of the conserved AHL. T3SS and non-T3SS secretins are highlighted green and orange, respectively. MxiD, *S. flexneri*; YscC, *Y. enterocolitica*; PscC, *P. aeruginosa*; EscC, EPEC; SpiA, *Salmonella* SPI-2; GspD, *Vibrio cholerae* T2SS; PulD, *Klebsiella oxytoca* T2SS; pIV, Phage F1; PilQ, *Neisseria meningitidis* type IV pilus system. Highly conserved residues are labelled and shown in **c**, including PulD multimerization mutants Ile569Ser (InvG Ile493 equivalent; solid square) and Phe573Leu (InvG Phe497 equivalent; solid star). **e**, InvG complementation assay for membrane-association mutants. Single alanine mutants in the AHL had no effect on secretion, Phe486Ala and Pro491Ala are shown as examples. Mutation of multiple conserved and hydrophobic residues in the AHL reduced secretion substantially, as did mutation of residues in the transmembrane β lip (Val337 and Leu339; labelled in **c**). Lysate InvG protein levels using anti-InvG antibody are indicated below the main panel.

the invaginated nature of the outer membrane in recent *in situ* tomography studies of various T3SSs^{20,25} (Extended Data Fig. 9d). Decorating the transmembrane lip in our secretin structure is an amphipathic helical loop (AHL), the most conserved sequence in the secretin family, presenting its hydrophobic face towards the inner leaflet of the

outer membrane (Fig. 3a, c, d). Amphipathic helices are commonly exploited for membrane association; examples include monotopic membrane proteins and peptide antibiotics, in which the hydrophobic side chains attach and partially penetrate into a single leaflet of the lipid bilayer. Mutants of the transmembrane β -lip and AHL still localize to the membrane (the AHL mutant to a lesser extent), a process that is probably mediated by the pilotin. Their ability to form SDS-resistant oligomers was, however, severely affected, indicating that they had a disruptive effect on membrane association and on insertion of the secretin pore²⁴ (Extended Data Fig. 9a), with a corresponding reduction in secretion (Fig. 3e). PulD mutants also map to the AHL (Ile569Ser and Phe573Leu; equivalent residues Ile493 and Phe497 of InvG; Fig. 3d and Extended Data Figs 5d, 6a) and were shown to affect the mislocalized insertion in the inner membrane¹⁵, multimerization kinetics¹⁵ and permeability (for PulD(Phe573Leu))²⁶. Taking these data together, we propose a role for AHL-mediated monotopic protein interactions with the outer membrane, in which they act as anchors to localize secretin monomers and facilitate assembly and oligomerization of the intimately associated N3 ring and secretin domain β -strands of the pre-pore state. This in turn provides a stabilized β -strand alignment of the secretin domain for the final 'zipping up' of the extensively hydrophobic β -barrel lip, a process that is required for full membrane insertion and span of the assembled, hyperstable pore (Extended Data Fig. 9b). The AHL could further affect localized physical membrane properties (as shown for other amphitropic proteins²⁷), such as fluidity or localized curvature. These membrane properties have been purported to be involved in PulD membrane insertion^{19,24}.

A prominent feature of secretins, as imaged by electron microscopy, is the periplasmic gate, a barrier necessary for ensuring selective passage of solutes. The gate in our structure is formed by the radial projections of β -strands 4 and 5 of the inner β -sheet, which extend into the secretin lumen creating a central pore approximately 15 Å across (Fig. 2b, c), consistent with the small-molecule permeability (less than ~600 Da) of secretins from the T2SS, T3SS and type IV pilus system^{26,28,29}. The pore is lined with the conserved residues Glu396 and Arg397, creating a ring of opposing charge on the extracellular and periplasmic face, respectively. Mutation of these, or of surrounding hydrophobic residues involved in hairpin packing, affects secretion considerably (Extended Data Fig. 10b, c). Furthermore, phage secretin pIV point mutants that lead to increased permeability can be mapped to the inner β sheet and N3 domain²⁹ (Extended Data Fig. 10b), supporting the involvement of both in gate-opening. The T3SS-specific N3 hairpin insertion packs against the underside of the inner β -sheet and mutation of conserved Arg198 and Asp199 at the hairpin tip reduces or eliminates secretion (Extended Data Fig. 10b, c). These residues appear to support the closed gate, with Arg198 forming a salt bridge with residues at the kink of the gate hairpin and Asp199 projecting up towards His403 on the underside of the gate. Taken together, we propose that the N3 domain has a role in gating by stabilizing the periplasmic gate in a closed position.

Comparison of *Salmonella* SPI-1 cryo-EM reconstructions^{8,30} shows that the assembly of the injectisome needle filament involves an outward reorientation of the region that we can now attribute to the N3 domain, which, in our closed basal body, is tilted into the secretin lumen (Extended Data Fig. 10a). Substrate-induced gate opening that occurs through interaction with the N3 domain has been proposed for the T2SS, on the basis of low-resolution electron microscopy³¹. Analogously to the type IV pilus system, in which the pilus (with similar complementary span and characteristics to the SPI-1 needle structure³²) passes through the secretin PilQ³³, rod/needle assembly within the injectisome secretin lumen would result in contact with the tightly packed N3 ring (Extended Data Fig. 10a, d–f). We envisage that this substrate-induced pushing of the N3 domain and the concomitant disruption of the N3 domain interface with the secretin domain inner β -sheet could provide the trigger for reorientation of the periplasmic gate strands and assembly of the conduit in these various secretion systems.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 February; accepted 25 October 2016.

Published online 14 December 2016.

- Worrall, L. J., Lameignere, E. & Strynadka, N. C. Structural overview of the bacterial injectisome. *Curr. Opin. Microbiol.* **14**, 3–8 (2011).
- Bergeron, J. R. *et al.* A refined model of the prototypical *Salmonella* SPI-1 T3SS basal body reveals the molecular basis for its assembly. *PLoS Pathog.* **9**, e1003307 (2013).
- Bergeron, J. R. *et al.* The modular structure of the inner-membrane ring component PrgK facilitates assembly of the type III secretion system basal body. *Structure* **23**, 161–172 (2015).
- Spreter, T. *et al.* A conserved structural motif mediates formation of the periplasmic rings in the type III secretion system. *Nat. Struct. Mol. Biol.* **16**, 468–476 (2009).
- Yip, C. K. *et al.* Structural characterization of the molecular platform for type III secretion system assembly. *Nature* **435**, 702–707 (2005).
- Korotkov, K. V., Gonen, T. & Hol, W. G. Secretins: dynamic channels for protein transport across membranes. *Trends Biochem. Sci.* **36**, 433–443 (2011).
- Schraidt, O. *et al.* Topology and organization of the *Salmonella* Typhimurium type III secretion needle complex components. *PLoS Pathog.* **6**, e1000824 (2010).
- Schraidt, O. & Marlovits, T. C. Three-dimensional model of *Salmonella*'s needle complex at subnanometer resolution. *Science* **331**, 1192–1195 (2011).
- Sanowar, S. *et al.* Interactions of the transmembrane polymeric rings of the *Salmonella enterica* serovar Typhimurium type III secretion system. *MBio* **1**, (2010).
- Korotkov, K. V., Pardon, E., Steyaert, J. & Hol, W. G. Crystal structure of the N-terminal domain of the secretin GspD from ETEC determined with the assistance of a nanobody. *Structure* **17**, 255–265 (2009).
- Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
- Linderoth, N. A., Model, P. & Russel, M. Essential role of a sodium dodecyl sulfate-resistant protein IV multimer in assembly-export of filamentous phage. *J. Bacteriol.* **178**, 1962–1970 (1996).
- Hardie, K. R., Lory, S. & Pugsley, A. P. Insertion of an outer membrane protein in *Escherichia coli* requires a chaperone-like protein. *EMBO J.* **15**, 978–988 (1996).
- Guilvout, I. *et al.* In vitro multimerization and membrane insertion of bacterial outer membrane secretin PulD. *J. Mol. Biol.* **382**, 13–23 (2008).
- Guilvout, I., Nickerson, N. N., Chami, M. & Pugsley, A. P. Multimerization-defective variants of dodecameric secretin PulD. *Res. Microbiol.* **162**, 180–190 (2011).
- Koo, J., Burrows, L. L. & Howell, P. L. Decoding the roles of pilotins and accessory proteins in secretin escort services. *FEMS Microbiol. Lett.* **328**, 1–12 (2012).
- Okon, M. *et al.* Structural characterization of the type-III pilot-secretin complex from *Shigella flexneri*. *Structure* **16**, 1544–1554 (2008).
- Gu, S., Rehman, S., Wang, X., Shevchik, V. E. & Pickersgill, R. W. Structural and functional insights into the pilotin-secretin complex of the type II secretion system. *PLoS Pathog.* **8**, e1002531 (2012).
- Huysmans, G. H., Guilvout, I., Chami, M., Nickerson, N. N. & Pugsley, A. P. Lipids assist the membrane insertion of a BAM-independent outer membrane protein. *Sci. Rep.* **5**, 15068 (2015).
- Hu, B. *et al.* Visualization of the type III secretion sorting platform of *Shigella flexneri*. *Proc. Natl Acad. Sci. USA* **112**, 1047–1052 (2015).
- Hoang, H. H. *et al.* Outer membrane targeting of *Pseudomonas aeruginosa* proteins shows variable dependence on the components of Bam and Lol machineries. *MBio* **2**, (2011).
- Dunstan, R. A. *et al.* Assembly of the secretion pores GspD, Wza and CsgG into bacterial outer membranes does not require the Omp85 proteins BamA or TamA. *Mol. Microbiol.* **97**, 616–629 (2015).
- Guilvout, I. *et al.* Independent domain assembly in a trapped folding intermediate of multimeric outer membrane secretins. *Structure* **22**, 582–589 (2014).
- Huysmans, G. H., Guilvout, I. & Pugsley, A. P. Sequential steps in the assembly of the multimeric outer membrane secretin PulD. *J. Biol. Chem.* **288**, 30700–30707 (2013).
- Nans, A., Kudryashev, M., Saibil, H. R. & Hayward, R. D. Structure of a bacterial type III secretion system in contact with a host membrane in situ. *Nat. Commun.* **6**, 10114 (2015).
- Disconzi, E. *et al.* Bacterial secretins form constitutively open pores akin to general porins. *J. Bacteriol.* **196**, 121–128 (2014).
- Papahadjopoulos, D., Moscarello, M., Eylar, E. H. & Isac, T. Effects of proteins on thermotropic phase transitions of phospholipid membranes. *Biochim. Biophys. Acta* **401**, 317–335 (1975).
- Burghout, P. *et al.* Structure and electrophysiological properties of the YscC secretin from the type III secretion system of *Yersinia enterocolitica*. *J. Bacteriol.* **186**, 4645–4654 (2004).
- Spagnuolo, J. *et al.* Identification of the gate regions in the primary structure of the secretin pIV. *Mol. Microbiol.* **76**, 133–150 (2010).

30. Marlovits, T. C. *et al.* Structural insights into the assembly of the type III secretion needle complex. *Science* **306**, 1040–1042 (2004).
31. Reichow, S. L., Korotkov, K. V., Hol, W. G. & Gonen, T. Structure of the cholera toxin secretion channel in its closed state. *Nat. Struct. Mol. Biol.* **17**, 1226–1232 (2010).
32. Loquet, A. *et al.* Atomic model of the type III secretion system needle. *Nature* **486**, 276–279 (2012).
33. Chang, Y. W. *et al.* Architecture of the type IVa pilus machine. *Science* **351**, aad2001 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Cheung for assistance with the expression trials of the PrgH_{130–392} GFP-fused basal body, C. Lizak for advice on GFP variants and membrane-protein purification and C. Yip and J. Rubenstein for advice on negative-stain TEM techniques. We thank UBC Bioimaging for access to TEM infrastructure. We thank K.-M. Moon and J. Rogalski at the Michael Smith Labs Proteomics Core Facility for assistance with LC–MS/MS. We thank F. Rosell at the LMB Spectroscopy and Kinetics Hub for assistance with circular dichroism. We thank S. Miller for providing the *S. Typhimurium* deletion strains and plasmids, as well as the InvG antibody. This work was funded by scholarships to L.W. and J.B. from the Canadian Institutes of Health Research (CIHR) and Michael Smith Foundation of Health Research, respectively, and operating grants from CIHR to N.C.J.S. and B.B.F., and the

Howard Hughes International Senior Scholar program to N.C.J.S. N.C.J.S. is a Tier I Canada Research Chair in Antibiotic Discovery.

Author Contributions L.J.W. performed all model building, refinement, structural analysis and modelling experiments. C.H. performed single particle cryo-EM grid preparation, data collection and map generation with input from R.K.H. and Z.Y. M.V. performed all cloning, basal body and secretin sample preparations used in the structure solution with input from L.J.W. and J.R.C.B., building on work initiated by T.S. L.J.W. and J.R.C.B. carried out negative-stain TEM analysis on basal body and secretin preps for quality control. L.J.W. and M.V. designed and made basal body mutants for secretion assays carried out by W.D. W.D. also generated/validated the *invG* deletion mutant. D.D.M. performed experiments probing isolated pilotin InvH and InvG S domain interactions. L.J.W. and N.C.J.S. principally wrote the manuscript with input from all.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.C.J.S. (ncjs@mail.ubc.ca) or Z.Y. (yuz@janelia.hhmi.org).

Reviewer Information *Nature* thanks M. Beeby, A. Blocker, J. Rubinstein and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Basal body purification. For basal body purification, a non-flagellated $\Delta prgH$ *Salmonella* strain³⁴ was used, complemented with a thrombin-cleavable N-terminal 6 \times His-GFP-tagged PrgH130–392 plasmid based on the *prgH* deletion plasmids described previously³⁴. The PrgH130–392-complemented *Salmonella* strain was transformed with a plasmid expressing the T3SS transcription activator HilA on a pBAD plasmid. Bacteria were grown with mild aeration in LB broth supplemented with 0.3 M NaCl to an A_{600} of approximately 0.5, arabinose was added to a final concentration of 0.015% and grown overnight at 30 °C. Cells were resuspended in 0.1 M Tris (pH 8.0), 0.75 M sucrose in the presence of lysozyme (0.4 mg ml⁻¹) and EDTA (1 mM), and incubated with stirring. Cells were lysed with 1% lauryldimethylamine oxide (LDAO) in the presence of protease inhibitor tablets (Roche) before adding 10 mM MgCl₂ and 500 mM NaCl to the lysate. Cell debris was removed by low-speed centrifugation and basal bodies were pelleted by high-speed centrifugation (Beckman, 45Ti rotor, 40,000 r.p.m., 45 min, 4 °C). The pellet was resuspended in 0.5% LDAO in 10 mM Tris-HCl pH 8, 0.5 M NaCl, 5 mM EDTA and adjusted to a final concentration of 30% (w/v) of CsCl. Following this, 5-ml samples were centrifuged for 16–20 h at 25,000 r.p.m. in a Beckman SW41 rotor. Fractions containing basal body were combined and desalted into 10 mM Tris (pH 8.0), 0.5 M NaCl, 0.2% LDAO before purification using nickel-charged Chelating Sepharose Fast Flow (GE Healthcare). Purified protein was concentrated and incubated with Thrombin overnight to cleave the N-terminal His-GFP tag. The sample was further concentrated before gel filtration on a Superose 6 column (GE Healthcare) equilibrated in 10 mM Tris (pH 8.0), 0.5 M NaCl, 0.2% LDAO.

To isolate the secretin, purified protein eluted from the nickel-charged Chelating Sepharose Fast Flow resin was incubated with 100 mM disodium hydrogen phosphate (pH 10.5) for 1 h, buffer was exchanged and passed back over the Chelating Sepharose Fast Flow resin before the final gel-filtration step.

Cryo-electron microscopy. Aliquots of 2.5 μ l of the purified basal body complex and the isolated secretin were applied to glow-discharged (30 s on each side) Quantifoil grids (Gold, 400 mesh, R1.2/1.3). The grids were blotted for 3 s at 80% humidity and plunge-frozen into liquid ethane using a VitroMark IV (FEI Company). Grids were imaged on a 300 keV Titan Krios cryo-electron microscope (FEI Company) equipped with a spherical aberration corrector, an energy filter (Gatan GIF Quantum) and a post-GIF Gatan K2 Summit direct electron detector. Images were taken on the K2 camera in dose-fractionation mode at a calibrated magnification of 29,240, corresponding to 1.71 Å per physical pixel (0.855 Å per super-resolution pixel). The dose rate on the specimen was set to be 3.4 electrons per Å² per s and total exposure time was 18 s, resulting in a total dose of 62 electrons per Å². With dose fractionation set at 0.375 s per frame, each movie series contained 48 frames and each frame received a dose of 1.3 electrons per Å². An energy slit with a width of 20 eV was used during data collection. Fully automated data collection was carried out using SerialEM with a nominal defocus range set from -1.5 to -3 μ m (ref. 35).

Image processing. For the basal body complex dataset, a total of 2,515 movie series were collected at super resolution (0.855 Å per pixel). Motion correction was done using Unblur³⁶, with data binned by two (1.71 Å per pixel after binning), and all 48 frames were aligned and summed to a single micrograph with dose-filtering using Sum_movie for further processing. The contrast transfer functions (CTFs) of the summed micrographs were determined using CTFFIND4 (ref. 37). Approximately 2,000 particles were manually boxed out from selected micrographs to generate an initial model using reference-free 2D-class averages by EMAN2 (ref. 38). The representative 2D-class averages were also used as templates for automated particle picking for the entire dataset in Relion³⁹. With ~263,900 particles automatically picked, reference-free 2D classification was performed in Relion and ~159,700 particles were kept in the good class averages. 3D classification in Relion generated one good 3D map out of four with ~67,800 particles. 3D refinement was performed with these 67,800 particles using FREALIGN⁴⁰, with and without imposing 24-fold symmetry. The refinement resolution upper limit was set to 8 Å in FREALIGN so that any information beyond 8 Å was not used in the refinement. Fourier shell correlations (FSCs) calculated from unmasked maps in FREALIGN at 0.143 criterion reported 6.3 Å resolution for the C1 map and 4.3 Å resolution for the C24 map using frequency limited refinement procedure. Final B-factor sharpening of the maps was performed using Bfactor.exe (Grigorieff laboratory). Local resolution estimations were performed using ResMap⁴¹ with the unfiltered and unmasked C1 map and C24 map, showing that certain regions exhibit better resolvability than the overall resolution.

For the InvG secretin pore dataset, a total of 2,685 movie series were collected at super resolution (0.855 Å per pixel). Motion correction, dose filtering and CTF correction were done in the same fashion as for the basal body dataset. Approximately 164,300 particles were initially picked using representative 2D class averages as templates in Relion³⁹. Relion was used for 2D classification and ~132,800 particles

were kept in good class averages. The initial 3D map of the secretin pore was cut from the corresponding part of the basal body initial map and low-pass filtered to 50 Å resolution. Relion 3D classification generated one good 3D map out of four, with ~83,900 particles. Relion was used to perform 3D auto-refinement, both with and without imposing C15 symmetry. Semi-automated post-processing of the maps, including automated soft masking, modulation transfer function and B-factor sharpening, was performed in Relion and yielded the final maps. FSCs at a criteria of 0.143 reported a 4.4 Å resolution for the C1 map and 3.6 Å resolution for the C15 map, using gold-standard refinement procedures. These FSC curves have been corrected for the soft-mask effects with high-resolution noise substitution⁴².

Model building, refinement and modelling. For the inner-membrane rings, the X-ray crystal structures of PrgH (PDB accession number 4G11), PrgK_{19–82} (PDB accession number 4W4M) and PrgK_{96–183} (PDB accession number 4OYC) were docked into the 24-fold averaged basal-body reconstruction with Chimera⁴³ and refined using density-guided iterative local refinement as implemented in Rosetta⁴⁴, taking into account the ring symmetry at all steps. Gaps in the model were manually built using Coot⁴⁵ where permitted by the density or by RosettaCM⁴⁶, guided by the experimental data, in less well-resolved regions. Rosetta iterative local refinement and Phenix⁴⁷ real-space refinement was carried out on the complete models (PrgH_{171–364} and PrgK_{20–203}) and Molprobity⁴⁸ and EMRinger⁴⁹ were used for validation, with the final models having good stereochemistry (Molprobity score 1.15, Ramachandran-favoured 94.39%, outliers 1.07%) and an EMRinger score of 2.1, typical of refined electron microscopy models in the 3–3.5 Å resolution range. For InvG, the density was of sufficient quality to permit manual tracing of most of the Co α -backbone in Coot and subsequent sequence assignment. Model refinement and building of regions with less well-resolved density proceeded as above, taking into account the 15-fold symmetry, resulting in a final model covering residues 172–557 (residues 1–171, 217–266 and 558–562 disordered) with a Molprobity score of 1.34 (Ramachandran-favoured 94.28%, outliers 2.41%) and EMRinger score of 3.1.

Rosetta electron-microscopy-guided symmetrical docking for the unresolved InvG N0 and N1 domains was carried as previously described² using the InvG_{34–173} X-ray crystal structure (PDB accession number 4G08; truncated C terminus to residue 171) and the unaveraged basal body reconstruction as a restraint.

Evolutionary conservation analysis was carried out with ConSurf⁵⁰ using both T3SS and all secretin homologues for InvG. Surface electrostatics analysis was carried out using APBS⁵¹.

LC-MS/MS. In-gel digestion for the basal body sample was carried out as described⁵². Digested proteins were loaded onto a Bruker Impact II Q-ToF mass spectrometer. Peptide separation was carried out on a 50-cm in-house-packed 75- μ m C18 column by a Proxeon EasynLC UPLC system, using 120-min water:acetonitrile gradients. Eluted peptides were ionized in positive ion mode, collecting MS/MS spectra for the top 15 peaks >500 counts, with a 30-s dynamic exclusion list.

Resulting data files were loaded into MaxQuant v1.5.1.0 for analysis⁵³. Identification used a 0.006 Da MS tolerance and 40 p.p.m. MS/MS tolerance, and a *Salmonella*-specific protein database containing all annotated *Salmonella* proteins extracted from Uniprot.

Generation of an in-frame, non-polar *invG* deletion mutant in *Salmonella*.

The *sacB* gene-based allelic exchange method and the suicide vector pRE112 (ref. 54) were used to generate an in-frame deletion mutant of *invG* in *S. enterica* serovar Typhimurium strain SL1344, which is resistant to streptomycin. Two DNA fragments flanking upstream (0.9 kb) and downstream (1.1 kb) of the coding region of *invG* were generated by PCR using primer pairs InvG-1 (KpnI underlined) (5'-GCGGTACCAGTTTCATAATGATTGCATCAGG-3') and DinVG-R (NheI underlined) (5'-GCGCTAGCACTAGAATAACCAGGTGTA ACCAG-3'), or DinVG-F (NheI underlined) (5'-GCGCTAGCAATATTCTGAA GCCAAGCGGTGCC-3') and InvG-2 (SacI underlined) (5'-GCGAGCTCAGTGAA GAGGGTATGGCTTTAC-3'). The two PCR products were first cloned into pCR2.1-TOPO (Invitrogen) and verified by DNA sequencing. After digestion with KpnI/NheI or NheI/SacI, respectively, the two DNA fragments were gel-purified and cloned into KpnI/SacI-digested pRE112 in a three-way ligation, generating pRE112- Δ invG, which contains 0.9 kb upstream and 1.1 kb downstream flanking sequences of *invG* and the *invG* gene with an internal in-frame deletion from nucleotides 76 to 1,596 (about 90% of the coding region). The coding regions of the 25 N-terminal and the 31 C-terminal amino acid residues of *invG* are retained to avoid any polar effects on the expression of the downstream *invE* gene that has its start codon overlapping with the stop codon of *invG*. An NheI site was introduced into the deletion site. The suicide vector pRE112- Δ invG was transformed into *E. coli* strain SM10 λ pir by electroporation, and introduced into *Salmonella* strain SL1344 by conjugation. After sucrose selection, *Salmonella* colonies resistant to sucrose and streptomycin and sensitive to chloramphenicol were screened for *invG* deletion by PCR. The *invG* mutants were further verified by multiple rounds of PCR. The *invG* deletion mutant thus created showed abolished

SPI-1-T3SS-mediated protein secretion, and this secretion defect can be complemented by expressing InvG from a plasmid (pMSD238; ref 55). For complementation assays, site-directed mutagenesis was carried out using a QuikChange mutagenesis kit (Stratagene).

SPI-1 secretion assays. *Salmonella* strains were grown overnight in LB broth (supplemented with antibiotics if required) at 37°C with shaking at 225 r.p.m. The cultures were then diluted 1:100 into 4 ml of fresh LB, with appropriate antibiotics, and grown under the same conditions for 6 h to induce SPI-1 type III secretion. The cultures were then centrifuged at 16,100g for 10 min to pellet the bacteria and the bacterial pellet was re-suspended in SDS-PAGE sample buffer to generate whole cell lysates. The growth medium supernatant was collected and passed through a Millex-GV 0.22 µm filter unit (Millipore) to remove any remaining bacteria and the secreted proteins were precipitated with a final concentration of 10% (v/v) trichloroacetic acid. The secreted proteins were then collected by centrifugation at 16,100g for 30 min, the protein pellet was dried in air and dissolved in SDS-PAGE sample buffer, with residual trichloroacetic acid neutralized using 0.5 µl of saturated Tris. The amount of the sample buffer used to re-suspend the bacterial pellet or dissolve the precipitated proteins was normalized according to the A_{600} values of the cultures to ensure equal loading of the samples. The secreted proteins were run on 12% SDS-PAGE gels and stained with Coomassie blue G250 (see Supplementary Information panels a–d for original SDS-PAGE gels and anti-InvG western blots). We note that some mutant proteins are present at sub-wild-type levels, especially those of the AHL implicated in membrane association. We propose that this is a result of decreased stability owing to impaired membrane association and insertion, and oligomerization.

InvG localization assay. A 50-ml secretion assay culture was collected, washed in PBS, and resuspended in 1 ml of 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 3 mM MgCl₂, 1 mM CaCl₂, with protease inhibitors (Roche). The sample was lysed by sonication and centrifuged at 5,000g to remove unlysed cells and debris. Supernatants were centrifuged in a Beckman Optima TLX Ultracentrifuge for 60 min at 100,000g to pellet the membranes. The supernatant corresponding to the soluble protein was collected and the membrane pellet washed once in lysis buffer. The final pellet, corresponding to the membrane fraction, was resuspended in 1 ml lysis buffer supplemented with 1% SDS. See Supplementary Information panel e for original anti-InvG western blot.

Expression and purification of InvH_{27–147} and InvG_{520–562}. To verify the interaction between the pilotin protein InvH and the S domain of the secretin InvG, the proteins were first isolated for use in isothermal titration calorimetry (ITC). To avoid complications from stabilizing the full-length lipidated InvH, an N-terminally truncated construct InvH_{27–147} was purified. His₆-tagged InvH_{27–147} and His₁₀-tagged InvG_{520–562} (wild type and the Trp448Ala/Lys449Ala/Val552Ala mutant) were expressed with pET28a in *E. coli* BL21 (DE3) at 22°C. Cells were collected by centrifugation and resuspended in 4 ml lysis buffer (20 mM HEPES (pH 7.5), 500 mM NaCl, 15 mM imidazole, one protease inhibitor mixture tablet (Roche)) per gram of cell paste. The cells were lysed with an Avestin cell homogenizer and centrifuged at 40,000g for 30 min to pellet insoluble material. The supernatant was loaded onto a 1-ml HisTrap HP Ni Sepharose column (GE Life Sciences) and washed with 40 ml of wash buffer (20 mM HEPES (pH 7.5), 500 mM NaCl, 50 mM imidazole), followed by a 10-ml wash with 75 mM imidazole. InvH_{27–147} was eluted in 10–20 ml of 20 mM HEPES (pH 7.5), 500 mM NaCl, and 250 mM imidazole. Both the wild-type and mutant constructs of InvG_{520–562} were instead eluted at 500 mM imidazole. The eluted fractions were desalted into 20 mM HEPES pH 7.5, 500 mM NaCl, and the His-tag cleaved overnight at 4°C with 1:1,000 Bovine α -thrombin (HTI). The thrombin-treated sample was incubated with 500 µl of Ni-charged chelating sepharose FF (GE Life Sciences) at 4°C for 30 min to remove uncleaved protein. The sample was concentrated and loaded onto a Superdex 75 10/300 GL column (GE Life Sciences) equilibrated with 20 mM HEPES (pH 7.5), 500 mM NaCl. The protein eluted as a single peak, and protein containing fractions were pooled and concentrated to 20 mg ml^{−1} (InvH_{27–147}) and 5 mg ml^{−1} (wild-type and mutant InvG_{520–562}).

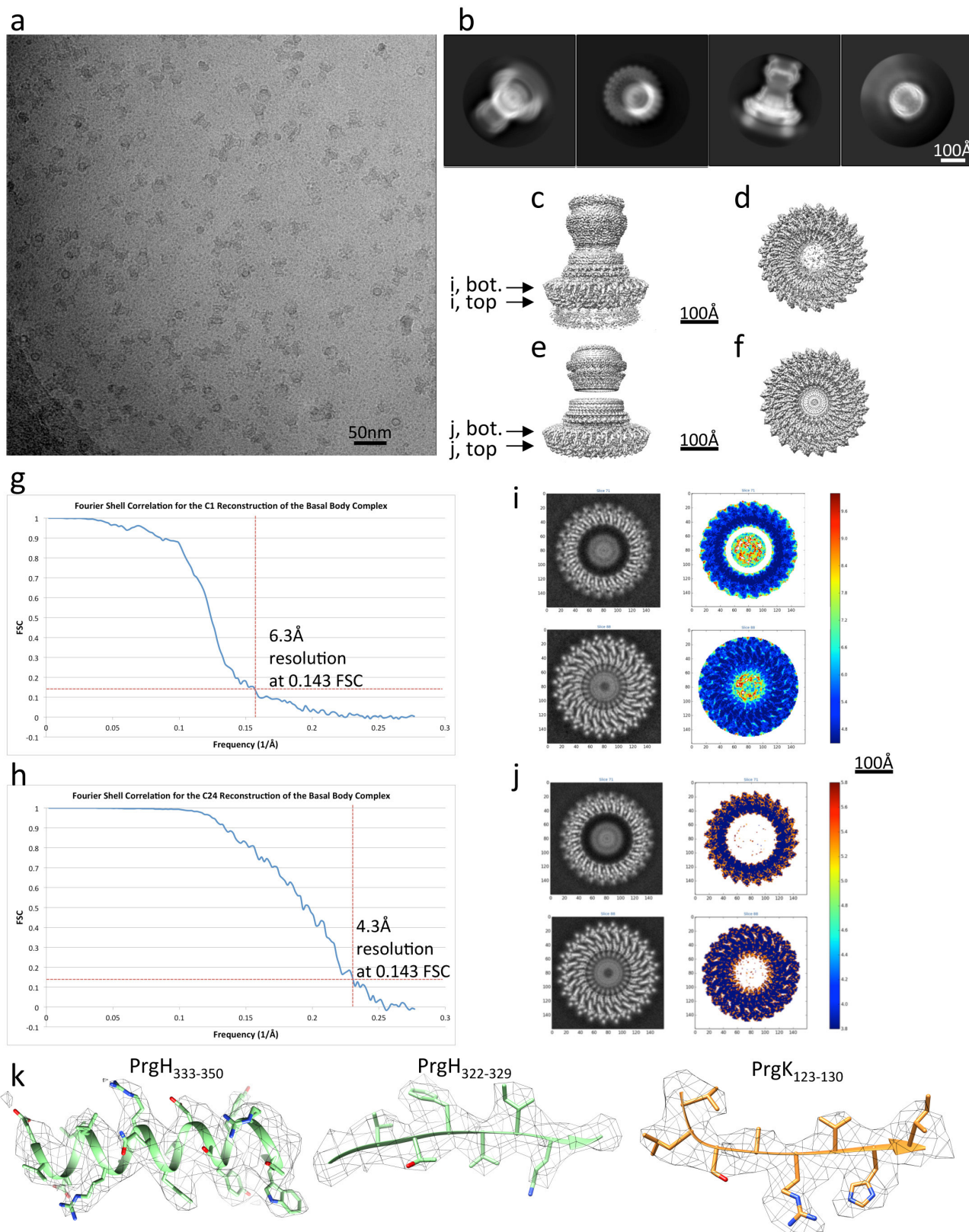
Isothermal titration calorimetry. ITC was performed using a MicroCal iTC200 (Malvern). All protein samples were dialysed overnight against 20 mM HEPES pH 7.5, 500 mM NaCl. The wild-type InvG_{520–562} (30 µM) contained in the ITC cell (200 µl volume) was titrated with twenty 2-µl injections of InvH_{27–147} (300 µM). The heat of dilution of the titrant was corrected for by subtracting a buffer control. Titrations of the Lys448Ala/Trp449Ala/Val552Ala triple mutant of InvG_{520–562} was performed in the same way, but at concentrations of 100 µM InvG_{520–562} in the cell and 1,000 µM InvH_{27–147} as the titrant to compensate for weaker binding. Four experiments were performed at 25°C. Data points were fitted to a one-binding-site model using the Origin 7 software (OriginLab Corporation). The algorithm was

used to optimize values for the binding stoichiometry (N), the association constant (K), and the binding enthalpy (ΔH), on the basis of the fit of the data.

Circular dichroism. Circular dichroism spectra of wild-type InvH_{27–147} and InvG_{520–562} were measured using a Jasco J-810 Circular Dichroism Spectrometer (Jasco Incorporated). Spectra of the individual proteins were measured at a concentration of 11 µM, whereas InvH_{27–147} and InvG_{520–562} were combined with a final concentration of 11 µM each to record the spectrum of the complex. Spectra were recorded in 25 mM Tris pH 7.5, 150 mM NaF at 25°C with a path length of 0.1 cm. Each spectrum represents the average of four scans, collected from 280 nm to 190 nm, with a spectral bandwidth of 1 nm and a response time of 2 s.

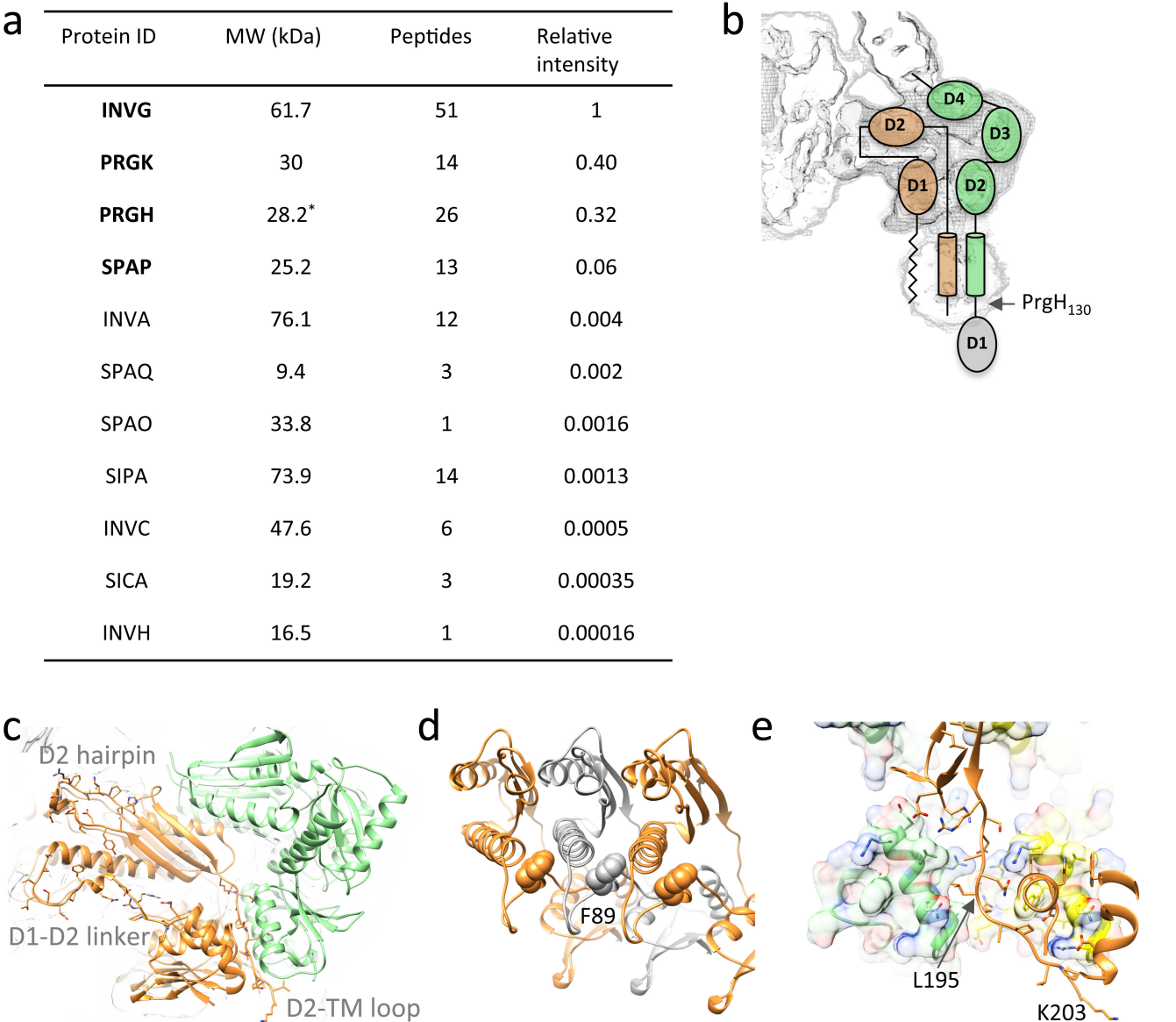
Data availability. Cryo-EM maps and atomic coordinates have been deposited with the Electron Microscopy Data Bank (with accession codes EMD-8398, EMD-8399, EMD-8400, EMD-8401) and Protein Data Bank (accession codes 5TCP, 5TCQ and 5TCR). The mass spectrometry data have been deposited with the ProteomeXchange Consortium via the PRIDE⁵⁹ partner repository with the dataset identifier PXD005133.

34. Kimbrough, T. G. & Miller, S. I. Contribution of *Salmonella* Typhimurium type III secretion components to needle complex formation. *Proc. Natl Acad. Sci. USA* **97**, 11008–11013 (2000).
35. Mastrorade, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
36. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).
37. Rohou, A. & Grigorieff, N. CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
38. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
39. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
40. Grigorieff, N. FREALIGN: high-resolution refinement of single particle structures. *J. Struct. Biol.* **157**, 117–125 (2007).
41. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
42. Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
43. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
44. DiMaio, F. *et al.* Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat. Methods* **12**, 361–365 (2015).
45. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
46. Song, Y. *et al.* High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735–1742 (2013).
47. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
48. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
49. Barad, B. A. *et al.* EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).
50. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, W529–W533 (2010).
51. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
52. Chan, Q. W., Howes, C. G. & Foster, L. J. Quantitative comparison of caste differences in honeybee hemolymph. *Mol. Cell. Proteomics* **5**, 2252–2262 (2006).
53. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
54. Edwards, R. A., Keller, L. H. & Schifferli, D. M. Improved allelic exchange vectors and their use to analyze 987P fimbria gene expression. *Gene* **207**, 149–157 (1998).
55. Daefler, S. & Russel, M. The *Salmonella* Typhimurium InvH protein is an outer membrane lipoprotein required for the proper localization of InvG. *Mol. Microbiol.* **28**, 1367–1380 (1998).
56. Hodgkinson, J. L. *et al.* Three-dimensional reconstruction of the *Shigella* T3SS transmembrane regions reveals 12-fold symmetry and novel features throughout. *Nat. Struct. Mol. Biol.* **16**, 477–485 (2009).
57. Kowal, J. *et al.* Structure of the dodecameric *Yersinia enterocolitica* secretin YscC and its trypsin-resistant core. *Structure* **21**, 2152–2161 (2013).
58. Tosi, T. *et al.* Structural similarity of secretins from type II and type III secretion systems. *Structure* **22**, 1348–1355 (2014).
59. Vizcaino, J. A. *et al.* The Proteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41**, D1063–D1069 (2013).



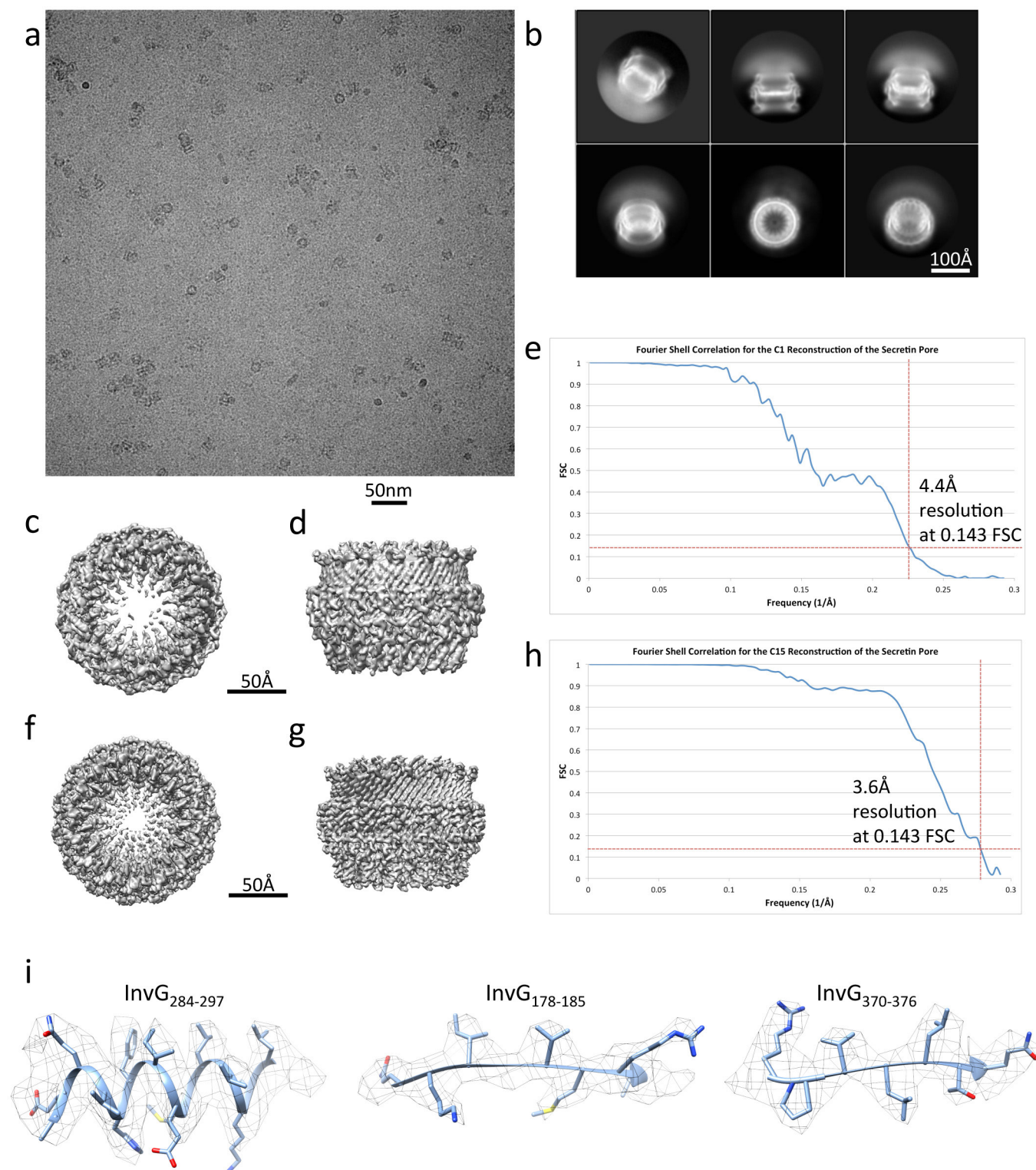
Extended Data Figure 1 | 3D reconstructions of the basal body complex. **a**, Representative micrograph of the basal body complex (a total of 2,515 were recorded). **b**, Selected reference-free 2D class averages. **c**, **d**, The side view (**c**) and the bottom view (**d**) of the C1 (no symmetry imposed) reconstructed map of the basal body complex. **e**, **f**, The side view (**e**) and the bottom view (**f**) of the C24 (24-fold symmetry imposed) reconstructed

map of the basal body complex. **g**, **h**, FSC of the C1 reconstruction (**g**) and the C24 reconstruction (**h**) calculated in FREALIGN using unmasked maps. **i**, Local resolution estimations of the C1 map. **j**, The C24 map from ResMap. Arrows in **c** and **e** indicate location of slice. **k**, Representative density for the inner-membrane rings (4.3 Å resolution).



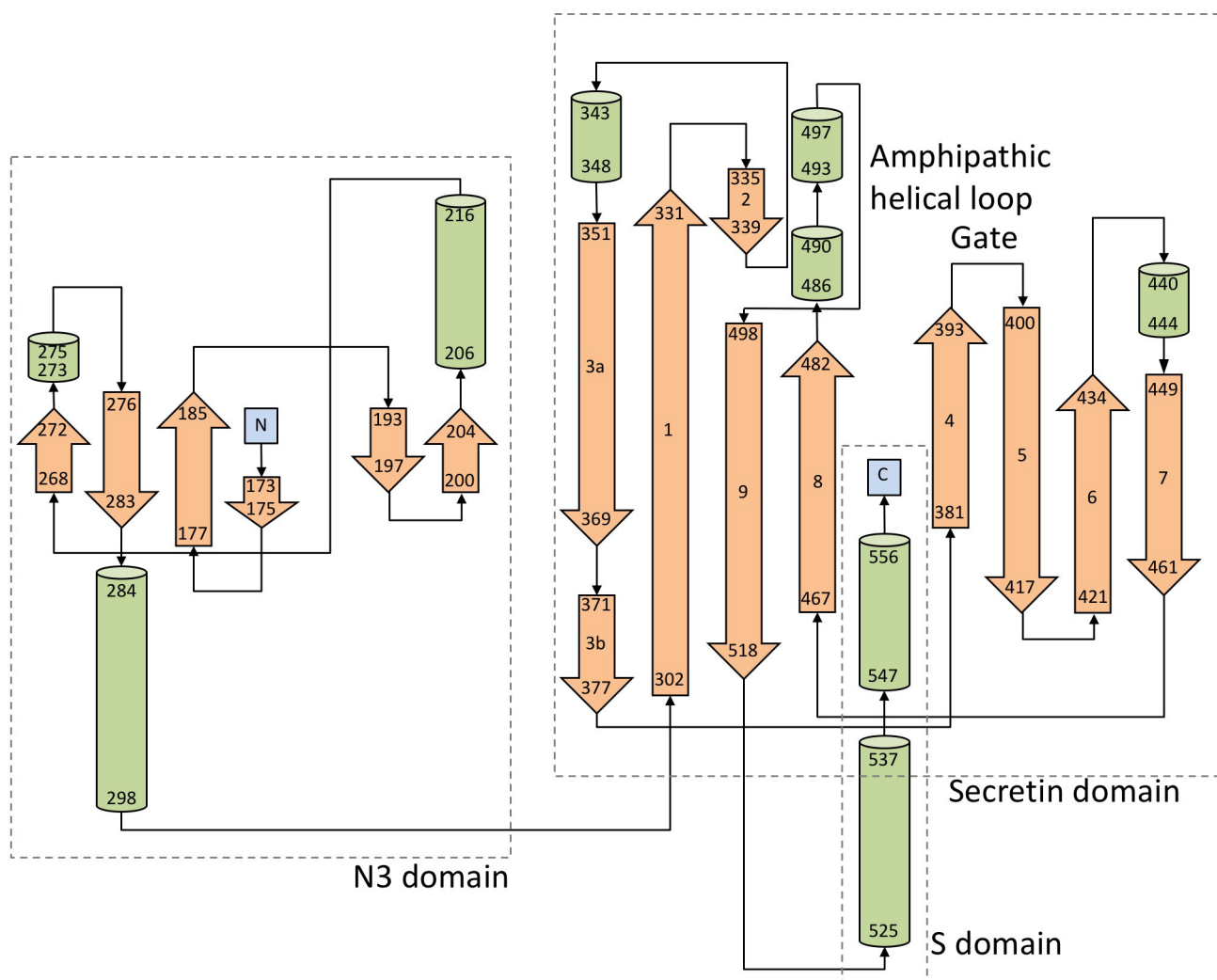
Extended Data Figure 2 | PrgH₁₃₀₋₃₉₂ basal body composition and oligomerization of PrgH and PrgK. **a**, Proteins from the purified PrgH₁₃₀₋₃₉₂ basal body identified by LC-MS/MS. InvG, PrgH, PrgK and SpaP were detected with elevated intensity. InvA and SpaQ peptides were also detected with decreased abundance. Some cytoplasmic export apparatus proteins and effectors were detected at very low levels, indicating they can still interact with the basal body complex. Only T3SS proteins are shown. *MW of PrgH₁₃₀₋₃₉₂. **b**, Schematic of PrgH and PrgK domain topology and position, as observed in our basal body structure. The PrgH cytoplasmic D1 domain is absent in the PrgH₁₃₀₋₃₉₂ mutant

and its location with respect to the basal body is unclear. **c**, Structure of PrgH₁₇₁₋₃₆₄ (green) and PrgK₂₀₋₂₀₃ (orange). Previously unresolved PrgK loops now observed in the assembled state are labelled and shown as sticks. **d**, Role of the PrgK D1-D2 linker in PrgK oligomerization. Residues 84-90 form a helix, with Phe89 (shown as spheres) inserting between neighbouring D2 domains. **e**, The D2 transmembrane loop contributes to oligomerization interacting with both the PrgK D1 domain and neighbouring PrgH protomers (green/yellow). Leu195 is the equivalent termination position in EPEC EscJ and the local environment of Lys203 was previously supported by chemical cross-linking⁹.

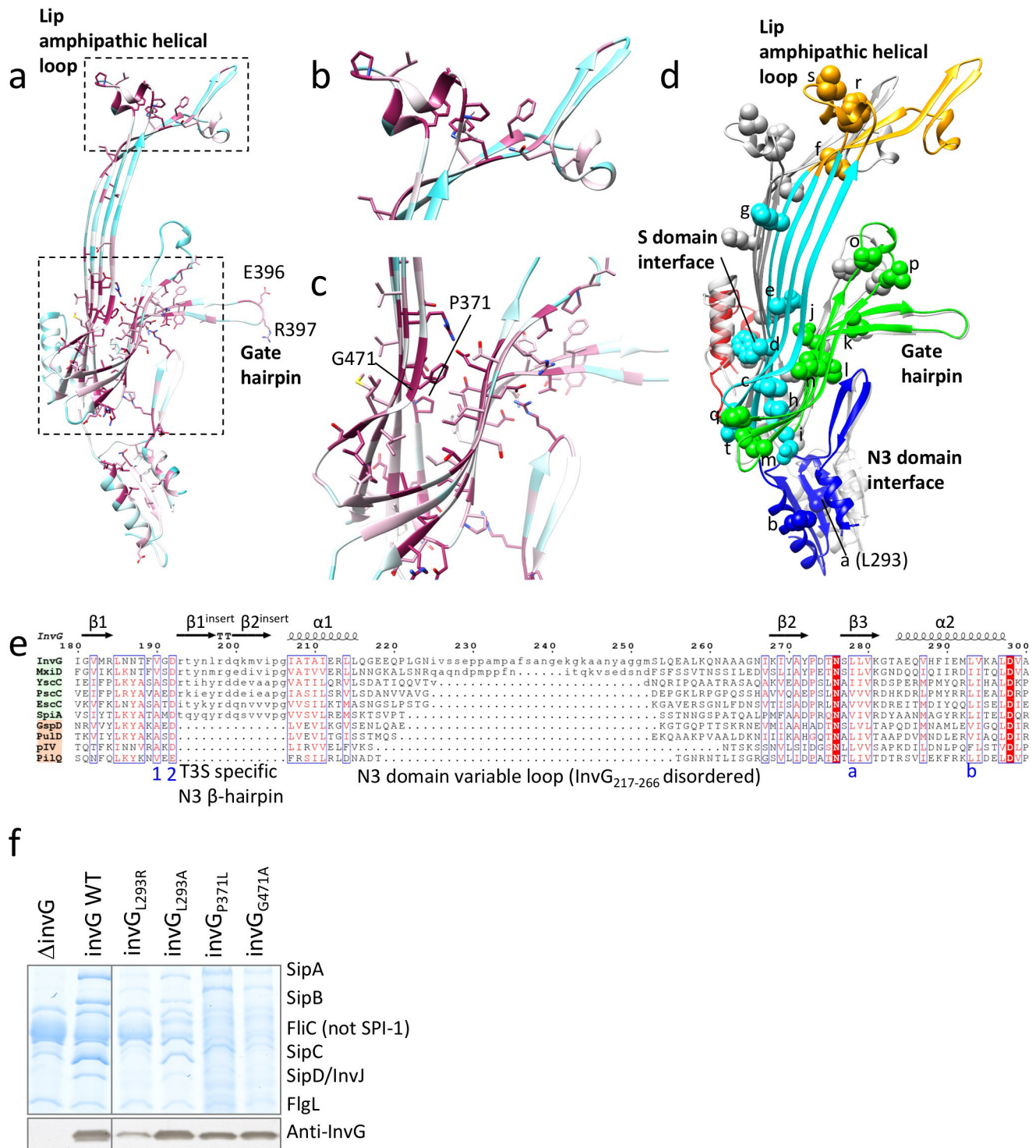


Extended Data Figure 3 | 3D reconstruction of the isolated secretin. **a**, Representative micrograph of the isolated secretin (total 2,685 recorded). **b**, Selected reference-free 2D class averages. **c, d**, The bottom view (**c**) and side view (**d**) of the C1 (no symmetry imposed) reconstructed map of the isolated secretin. **e**, FSC curve of the C1 reconstruction using gold-standard refinement with soft-masking-effect correction.

f, g, Bottom view (**f**) and side view (**g**) of the C15 (15-fold symmetry imposed) reconstructed map of the isolated secretin. **h**, FSC curve of the C15 reconstruction using gold-standard refinement with soft-masking-effect correction. **i**, Representative density for the isolated secretin (3.6 Å resolution).

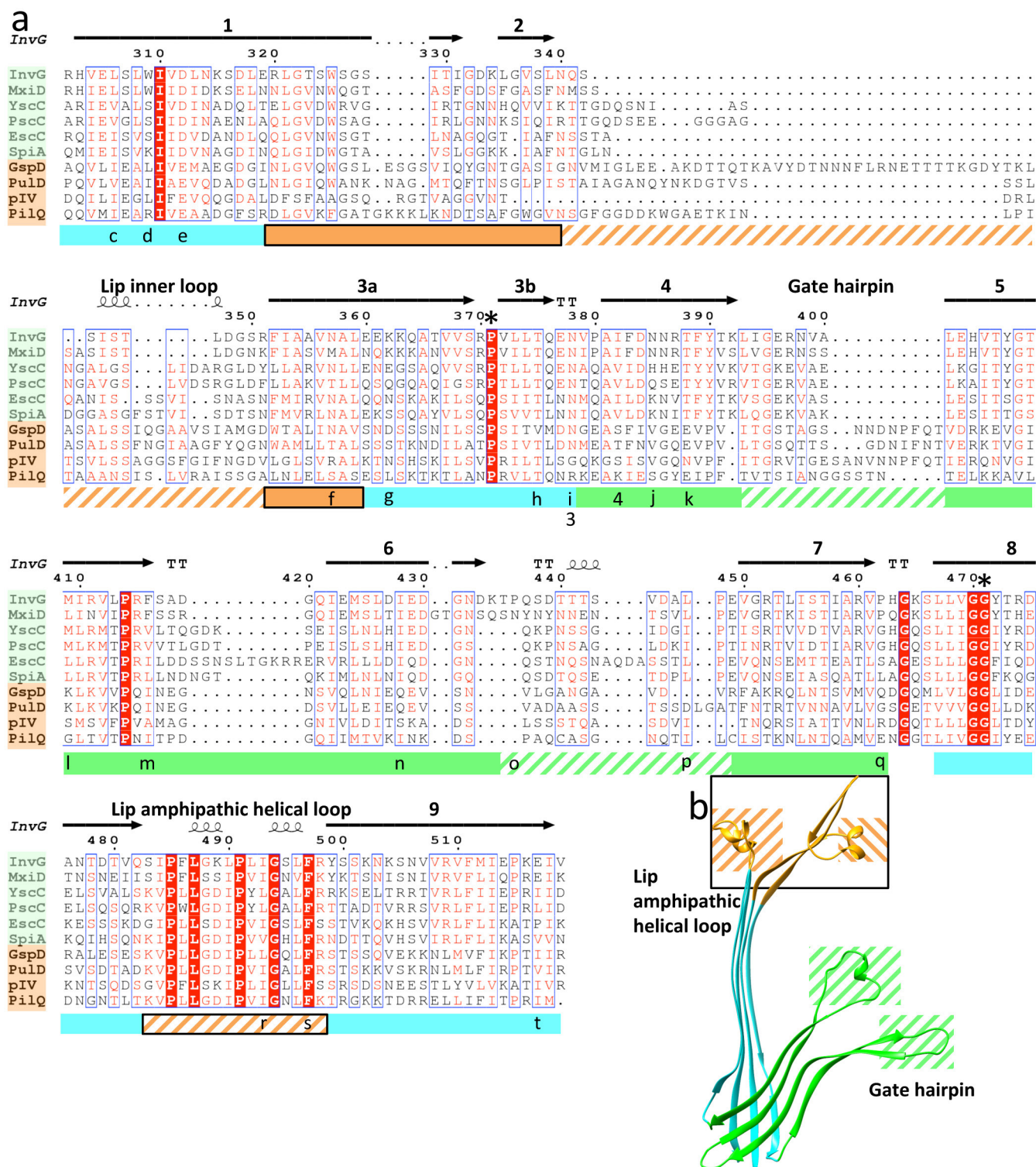


Extended Data Figure 4 | Secondary structure topology of the InvG secretin. Secondary structure topology for InvG_{172–557}. β -strands of the secretin domain are numbered, with 1, 3a/3b, 8 and 9 forming the outer β -barrel; 4–7 forming the inner β -barrel; and 1, 2 and 3a forming the lip of the β -barrel. Strand 3 is broken into 3a and 3b by the conserved residue Pro371.



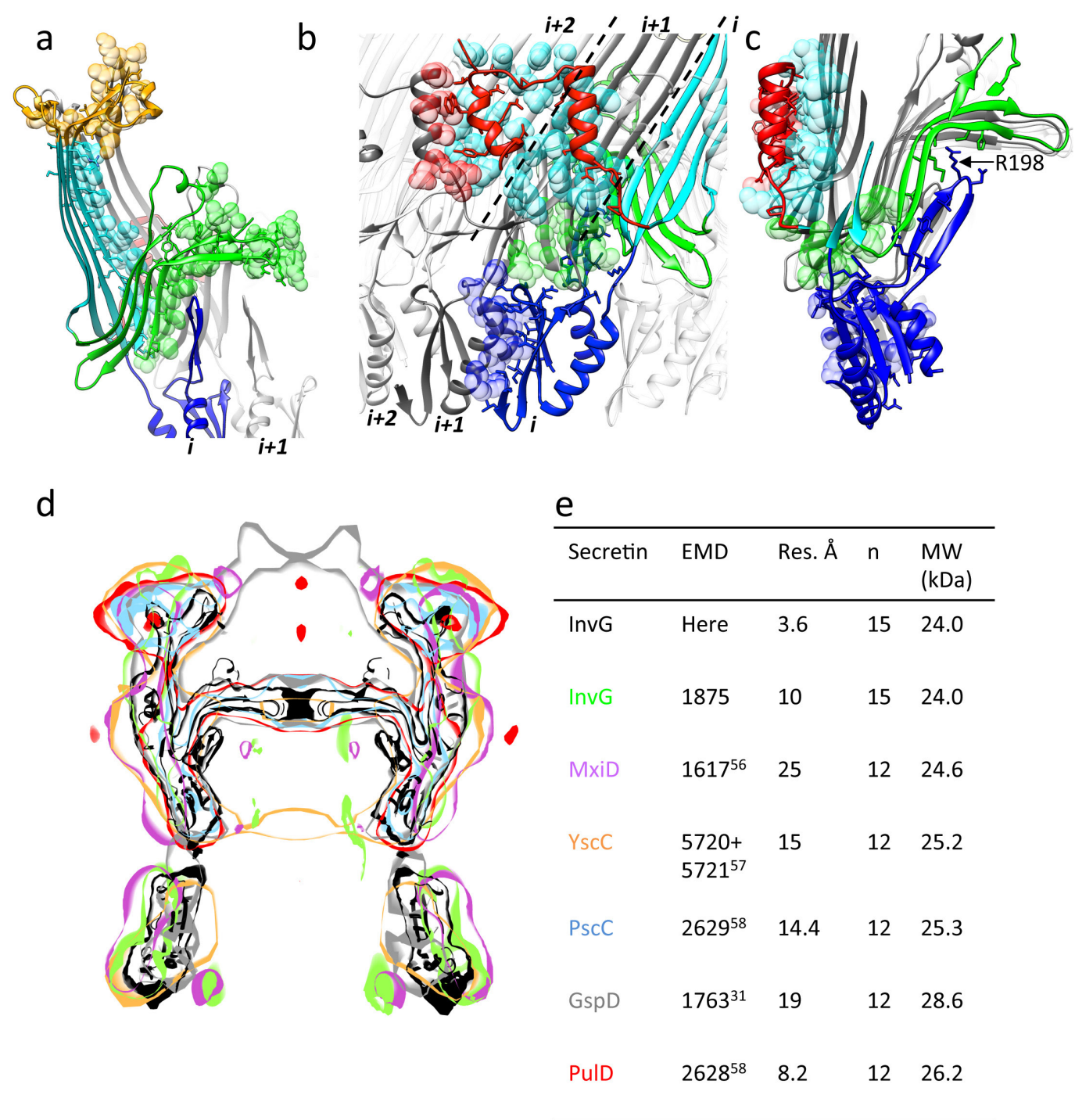
Extended Data Figure 5 | Sequence conservation of secretins and structure-based mapping of previously characterized PuID oligomerization mutants. **a**, InvG₁₇₂₋₅₅₇ sequence conservation (see Methods), coloured from magenta (highest) to cyan (lowest) for T3SS homologues. **b**, **c**, Close-up views of boxed regions in **a**. Pro371 and Gly471 (see **f**), highly conserved in all secretins (see Extended Data Fig. 6a), are labelled in **c**. **d**, PuID multimerization mutants mapped onto the InvG structure (shown as spheres). Coloured according to Fig. 2. Residues are labelled with letters and correspondingly annotated in **e** (N3 domain alignment; a–t) or Extended Data Fig. 6a (secretin domain alignment; c–t). **e**, Alignment of N3 domain (InvG₁₈₀₋₃₀₀) from T3SS (green) and

non-T3SS (orange) secretin homologues (see Fig. 3d). Conserved regions are lettered red and boxed. Invariant residues are boxed in solid red. Secondary structural elements observed here are annotated and numbered, consistent with other RBMs. PuID multimerization and pIV permeability mutants mapping to this domain are indicated by letters or numbers, respectively, and shown in **d** (for PuID) or Extended Data Fig. 10b (for pIV). **f**, InvG complementation assay for N3 domain ring-interface mutants (Leu293Arg, Leu293Ala; labelled in **d**) and conserved secretin domain β sandwich mutants (Pro371Leu, Gly471Ala; also labelled in **c** and Extended Data Fig. 6a). Lysate InvG protein levels indicated below using anti-InvG antibody.



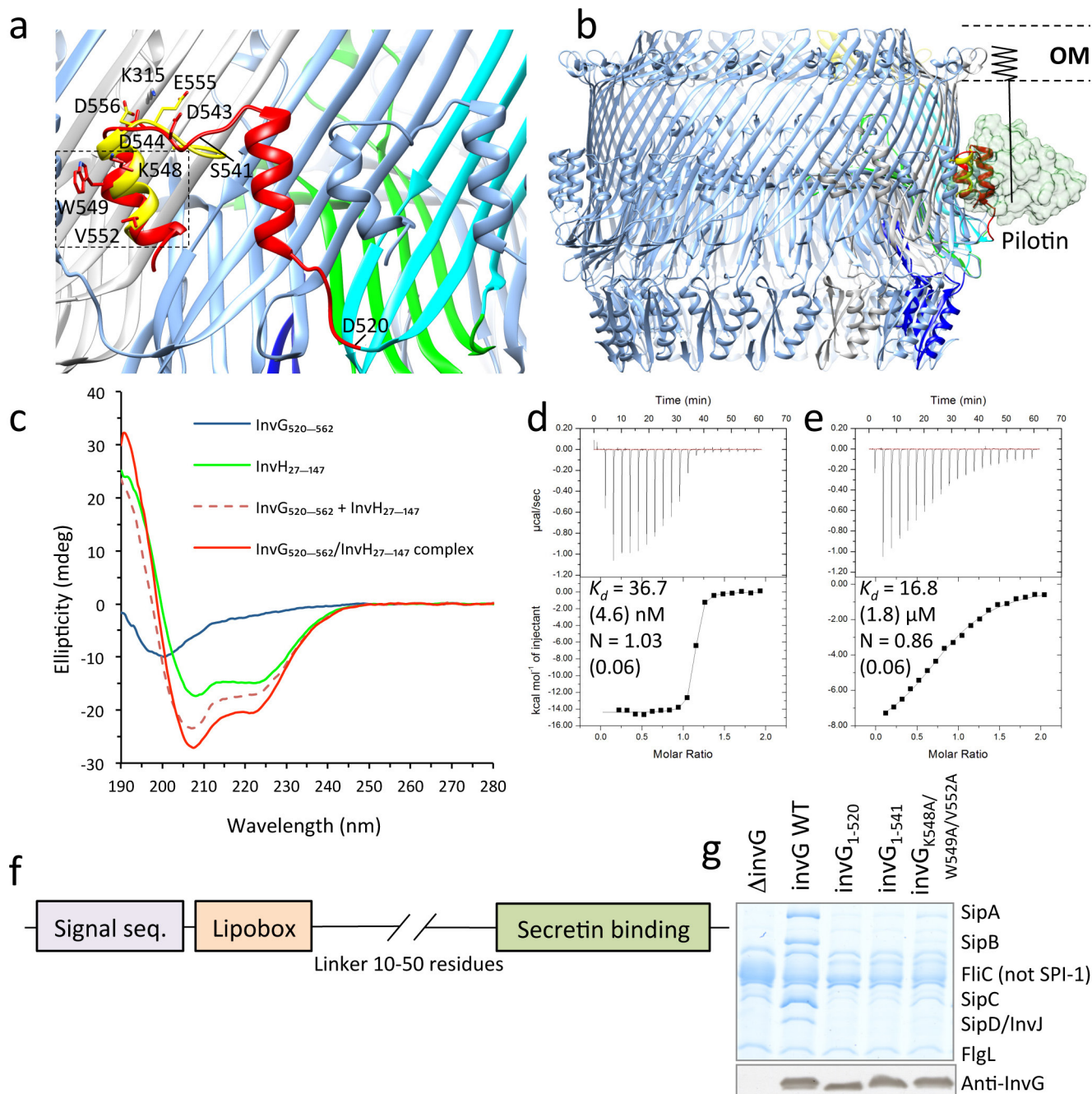
Extended Data Figure 6 | Sequence alignment for T3SS and non-T3SS secretin domains. **a**, Alignment of secretin domain (*InvG*_{302–519}) from T3SS (green) and non-T3SS (orange) secretin homologues (see Fig. 3d). Anti-parallel strands of the outer and inner β -sheets coloured cyan and green respectively with loop regions (where most indels are located) denoted by pattern of diagonal lines. Strands numbered according to Extended Data Fig. 4. Conserved residues Pro371 and Gly471

(Extended Data Fig. 5f) are marked with asterisks. *PulD* multimerization and *pIV* permeability mutants mapping to this domain are indicated by letters or numbers respectively and shown in Extended Data Fig. 5d (*PulD*) or Extended Data Fig. 10b (*pIV*). **b**, Corresponding structural elements. The residues that make up the putative membrane interacting region (gold) are boxed in both the structure and the sequence alignment.



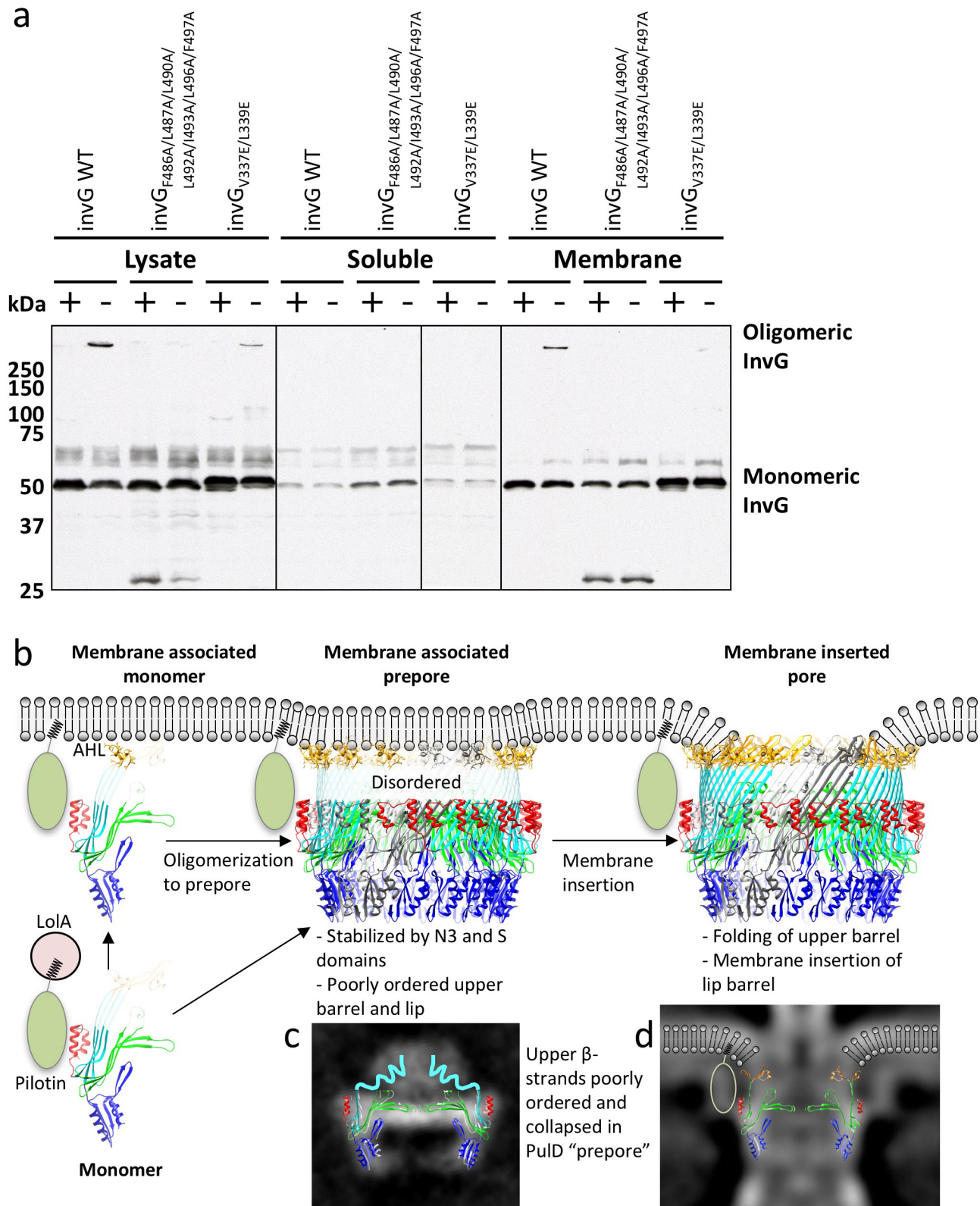
Extended Data Figure 7 | Secretin oligomerization and stoichiometry. **a**, Secretin domain interface between neighbouring monomers (denoted as *i* and *i* + 1). To differentiate adjacent neighbours and their extensive interfaces, the *i* ribbon is coloured as in Fig. 2 with the residues forming the oligomeric interfaces shown as sticks. The *i* + 1 ribbon is coloured grey with the residues at the oligomeric interface shown as transparent Van der Waal spheres and coloured according to domain feature. **b**, **c**, Role of N3 (cobalt blue) and S (red) domains in secretin multimerization. The *i* ribbon is coloured as in Fig. 2 with N3 and S domain interface residues shown as sticks, neighbouring monomers (*i* + 1 and *i* + 2) coloured grey and white, respectively (and also delineated by dashed line), with the residues interacting with N3 or S domains from monomer *i* displayed as spheres and coloured as in Fig. 2. The N3 domain of monomer *i*

(blue sticks) interfaces with both the N3 domain (blue spheres) and underside of the inner β sheet (green spheres) of monomer *i* + 1. The S domain of monomer *i* (red sticks) forms an extensive stapled interface with the outer β -sheet of *i* + 1 and *i* + 2 monomers (cyan spheres) and with the N-terminal S domain helix of monomer *i* + 1 (red spheres). **d**, **e**, Superimposition of available secretin electron microscopy maps and table showing EMD accession number, resolution, stoichiometry (*n*) and molecular weight (MW) of the secretin domain monomer. All secretins have the same general architecture, with GspD having an elongated upper lip and enclosed upper chamber as a result of the loop 1 insertion following strand 2 (Extended Data Fig. 6a). MxiD (*S. flexneri* T3SS), YscC (*Y. enterocolitica* T3SS), PscC (*P. aeruginosa* T3SS), GspD (*V. cholerae* T2SS), PulD (*K. oxytoca* T2SS).



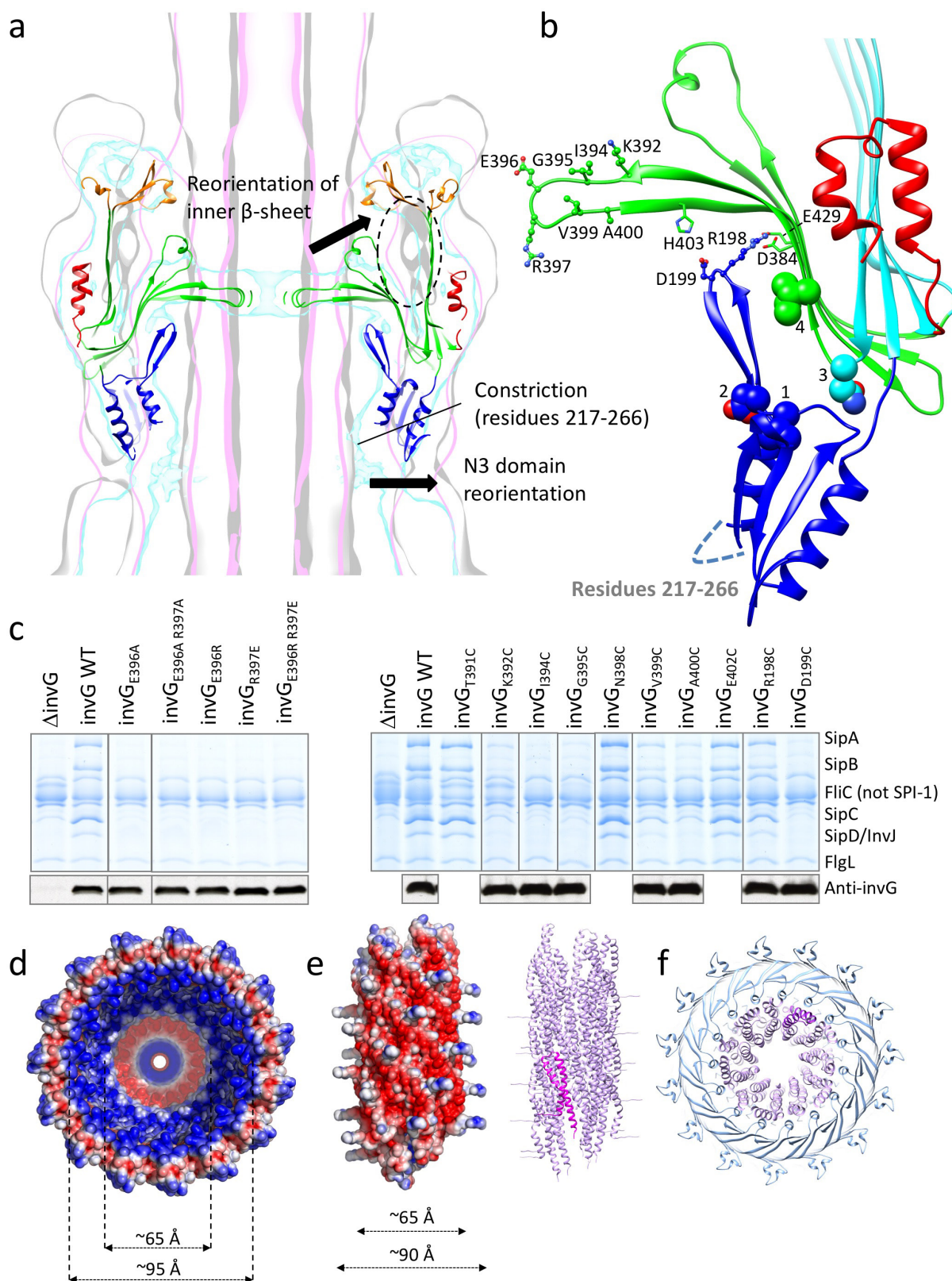
Extended Data Figure 8 | Interaction of the secretin S domain with the pilotin. **a**, InvG S domain (red) superposed with the MxiM/MxiD pilotin-bound secretin peptide (yellow) from *S. flexneri*. The electronegative MxiD Glu555 and Asp556 superpose with InvG Asp543 and Asp544. InvG Asp544 forms a salt bridge with Lys315, which is also conserved between MxiD and InvG (shown as sticks). Ordering of this helical peptide of the secretin S domain has been shown to occur upon binding to the pilotin. **b**, Superimposition of pilotin MxiM (green surface) onto InvG, based on binding to a common secretin peptide as in **a**, showing the S domain orientation is permissive of an assembled interaction with pilotin in the oligomerized form. Putative linker (solid line) and membrane inserted lipidation shown. OM, outer membrane. **c**, Far-UV circular dichroism spectra (from an average of four scans) for InvG_{520–562} (blue line), InvH_{27–147} (green line) and the complex (red line) showing an approximately

20% increase in ellipticity at 222 nm as compared to the calculated combined spectra (red dashed line), indicative of increased helical content in the complex and consistent with a disorder–order transition in the InvG S domain. **d**, **e**, ITC analysis for the interaction between InvH_{27–147} and InvG_{520–562} and InvG_{520–562}(K548A/W549A/V552A) respectively. A representative run is shown (from four runs) and K_d and N (stoichiometry) are reported as mean (standard deviation) from four runs. **f**, Generalized schematic of pilotin domains. N-terminal type II signal sequence followed by a lipobox lipidation signal with conserved cysteine connected to a structurally diverse globular C-terminal secretin binding domain via a variable length linker. **g**, InvG complementation assay for S domain deletion mutants and C-terminal helix triple mutant. Lysate InvG protein levels indicated below using anti-InvG antibody.



Extended Data Figure 9 | InvG membrane localization and proposed pathway for pore assembly and membrane insertion. **a**, Distribution of wild-type (WT) InvG and mutants in whole-cell lysate, soluble and membrane fractions. Ability to form SDS-resistant oligomer assessed by running both boiled (+) and unboiled (–) samples. The AHL mutant InvG(F486A/L487A/L490A/L492A/I493A/L496A/F497A) and transmembrane β -strand mutant InvG(V337G/L339G) (mutants that abrogate and reduce secretion, respectively; see Fig. 3e) both localize to the membrane (the AHL mutant to a lesser degree); however, their ability to form SDS-resistant oligomers is substantially affected and protease sensitivity is evident for the AHL mutant—both observations are consistent with aberrant membrane association, insertion and final stabilized assembly²⁴. **b–d**, Schematic of the proposed secretin assembly and membrane insertion pathway. The secretin monomer is localized to the outer membrane in a Lol-dependent manner, most likely in a complex with the pilotin, where we propose initial membrane

association is mediated by the conserved amphipathic loop (AHL; gold). Oligomerization to a membrane associated pre-pore intermediate follows which, based on our structure and earlier negative stain electron microscopy of this form in PulD (**c**; figure reproduced from ref. 23), would encompass a folded secretin core and the peripheral N3 and S domains with the extensive interfaces between them (see Extended Data Fig. 7a–c) stabilizing this pre-pore form. Subsequent folding of the remainder of the secretin β -domain (disordered in the pre-pore image) to create the upper β -barrel lip (gold) leads to the final, BAM-independent, membrane insertion and creation of the membrane-spanning secretin pore. Outer-membrane curvature observed *in situ* in several T3SSs is illustrated by overlaying the secretin structure with the *in situ* tomography of the *Shigella* injectisome (**d**; figure reproduced from ref. 20) also showing presence of continuous density, proposed to be the pilotin (circled), connecting the inner leaflet of the outer membrane and the region corresponding to the S domain in our InvG structure.



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | The secretin periplasmic gate.

a, Superimposition of the structure of InvG (coloured according to Fig. 2), basal body map (cyan), the approximately 10 Å needle-complex map (EMD-2481; grey) and the approximately 20 Å needle-complex map (EMD-1100; pink). Our basal body has dimensions with greater similarity to EMD-1100. Substrate passage through the secretin would require N3 domain reorientation and subsequent opening of periplasmic gate indicated with black arrows. Differences in the regions corresponding to the upper secretin β -sandwich in the needle-complex maps (circled) suggest a conformation of the open gate packed against the outer β -sheet. **b**, InvG structure, showing the periplasmic gate and N3 domain residues implicated in gating. InvG mutants with a secretion-deficient phenotype (shown in **c**) are shown as balls and sticks and the location of the

N3-secretin interface-localized phage secretin pIV permeability mutants (additional pIV permeability mutants are mapped throughout the inner β -sheet) are shown as spheres (labelled 1–4 and correspondingly annotated in Extended Data Fig. 5e (N3 domain alignment; 1–2) or Extended Data Fig. 6a (secretin domain alignment; 3–4)). **c**, InvG complementation assay for periplasmic-gate and N3-hairpin mutants. Lysate InvG protein levels are indicated below using anti-InvG antibodies. **d**, Electrostatic surface of the InvG pore structure, including the Rosetta-modelled N0 and N1 domains, as viewed from the periplasmic face. **e**, Electrostatic surface (left) of the *Salmonella* SPI-1 needle (PDB 2LPZ) and corresponding ribbon representation (right). A single PrgI monomer is coloured magenta. **f**, Superimposition of *Salmonella* SPI-1 needle (purple) onto the InvG_{172–557} pore structure (blue), as viewed from the extracellular face.

High-resolution crystal structure of the human CB1 cannabinoid receptor

Zhenhua Shao¹, Jie Yin¹, Karen Chapman¹, Magdalena Grzemska¹, Lindsay Clark¹, Junmei Wang² & Daniel M. Rosenbaum¹

The human cannabinoid G-protein-coupled receptors (GPCRs) CB1 and CB2 mediate the functional responses to the endocannabinoids anandamide and 2-arachidonyl glycerol (2-AG) and to the widely consumed plant phytocannabinoid Δ^9 -tetrahydrocannabinol (THC)¹. The cannabinoid receptors have been the targets of intensive drug discovery efforts, because modulation of these receptors has therapeutic potential to control pain², epilepsy³, obesity⁴, and other disorders. Although much progress in understanding the biophysical properties of GPCRs has recently been made, investigations of the molecular mechanisms of the cannabinoids and their receptors have lacked high-resolution structural data. Here we report the use of GPCR engineering and lipidic cubic phase crystallization to determine the structure of the human CB1 receptor bound to the inhibitor taranabant at 2.6-Å resolution. We found that the extracellular surface of CB1, including the highly conserved membrane-proximal N-terminal region, is distinct from those of other lipid-activated GPCRs, forming a critical part of the ligand-binding pocket. Docking studies further demonstrate how this same pocket may accommodate the cannabinoid agonist THC. Our CB1 structure provides an atomic framework for studying cannabinoid receptor function and will aid the design and optimization of therapeutic modulators of the endocannabinoid system.

The endocannabinoid signalling system in mammals comprises endogenous lipid messengers (anandamide and 2-AG) and two homologous GPCRs (CB1, which is located in the nervous system and periphery, and CB2, which is expressed primarily in immune cells)¹. Human CB1 and CB2 (which share 42% sequence identity) are also activated by natural products⁵ such as THC and by synthetic cannabinoids and can be inhibited by diverse subtype-selective and non-selective antagonists and inverse agonists⁶. CB1 is the most abundant GPCR in the central nervous system (CNS) and regulates diverse brain functions and behaviours, modulating neurotransmitter release and neuronal excitation through the pre-synaptic activation of the G-protein $G_{i/o}$ (inhibiting adenylate cyclase), GIRK channels, and arrestin/MAP kinase signalling⁷. Endocannabinoids are synthesized postsynaptically by lipases and travel across synapses in a retrograde manner⁸, embedding in the presynaptic membrane where they can activate CB1⁹. Beyond the CNS, CB1 signalling in peripheral tissues has been implicated in other physiological mechanisms such as release of the metabolic hormones leptin and insulin^{10,11}. However, the mechanism by which lipidic or lipophilic cannabinoid agonists access their GPCR-binding sites and promote receptor activation through specific binding interactions is, as yet, unknown.

Although humans have been consuming phytocannabinoids for their psychotropic effects for thousands of years¹, THC was only found to be the active chemical constituent of *Cannabis sativa* in 1964 (ref. 12). Recently, alternative therapeutic uses for cannabinoid ligands have been pursued. As the endocannabinoid system is involved in the regulation of energy metabolism⁴, synthetic inverse agonists such as rimonabant and

taranabant have proven effective in the clinic for treatment of obesity, but have failed to secure regulatory approval owing to adverse CNS side effects¹³. Peripheral blockade of CB1 by non-penetrant inverse agonists may represent an alternative therapeutic strategy for treating obesity, while avoiding CB1 receptors in the CNS¹⁰. Natural and synthetic cannabinoid ligands have also shown considerable promise in the treatment of neuropathic pain² and epilepsy-induced seizures³. To gain further insight into the molecular mechanisms of cannabinoid system modulators and aid structure-based ligand design, we sought to crystallize the human CB1 receptor and solve its atomic structure.

Obtaining diffraction-quality crystals of CB1 required optimization of both the construct and the purification technique. We carried out differential scanning fluorimetry on the detergent-solubilized receptor, which identified the inverse agonist taranabant as a ligand conferring enhanced thermostability (Methods and Extended Data Fig. 1). To promote lipidic cubic phase (LCP) crystallization, we replaced the third intracellular loop (ICL3) of CB1 with the thermostable PGS (*Pyrococcus abyssi* glycogen synthase) domain, which recently proved essential in helping solve crystal structures of the human orexin receptors¹⁴. We also incorporated the point mutation T210A, which was previously shown to stabilize the inactive conformation of CB1 and increase thermostability¹⁵. Finally, we truncated CB1(T210A)-PGS by eliminating the first 89 N-terminal residues and the C terminus after residue 421. The engineered construct binds to the inverse agonists taranabant and

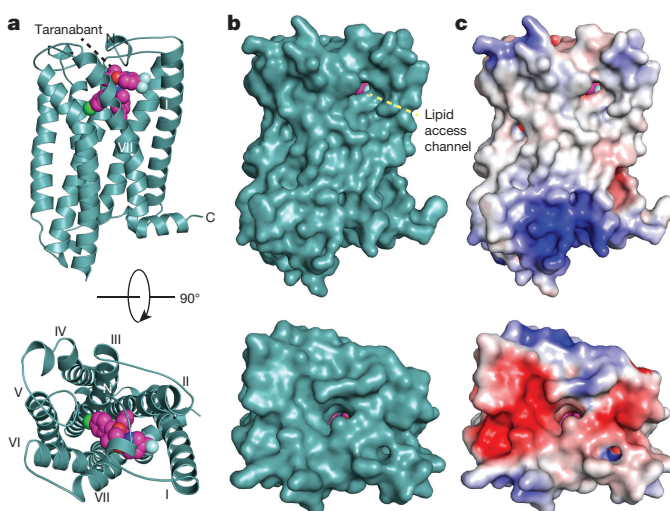


Figure 1 | Global structure of CB1 bound to taranabant. **a**, CB1 is represented as a teal cartoon. The taranabant ligand is shown as spheres with magenta carbon atoms. Views are from within the plane of the membrane (top) and from the extracellular space (bottom). **b**, Solvent-accessible surface representation of CB1 from the same views as in **a**. **c**, CB1 surface representation coloured according to electrostatic potential, from red (negative) to blue (positive).

¹Department of Biophysics, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ²Green Center for Systems Biology, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA.

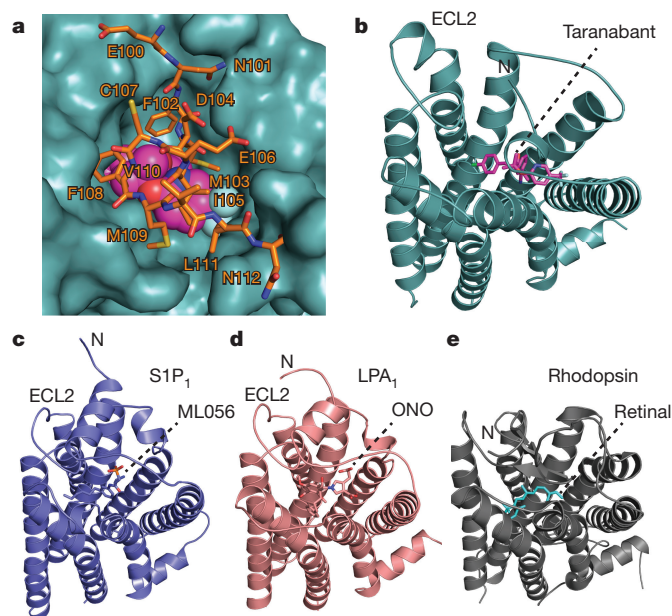


Figure 2 | Membrane-proximal N-terminal region of CB1. **a**, Interaction between the membrane-proximal N-terminal region and the rest of the receptor. Residues 100–112 are shown as orange sticks, and taranabant is shown as magenta spheres. The remainder of CB1 is depicted as a teal solvent-accessible surface. **b**, The extracellular region of CB1, with the receptor as a teal cartoon and taranabant as magenta sticks. **c**, The S1P₁ receptor (PDB accession number 3V2Y) is depicted as a blue cartoon, from the same perspective as in **a** after superposition with CB1. The ML056 antagonist is shown as blue sticks. **d**, The LPA₁ receptor (PDB accession number 4Z35) as a salmon cartoon, from the same perspective as in **a** after superposition with CB1. The ONO 9910539 antagonist is shown as salmon sticks. **e**, The GPCR rhodopsin (PDB accession number 1F88) as a grey cartoon, from the same perspective as in **a** after superposition with CB1. The 11-*cis*-retinal inverse agonist ligand is shown as cyan sticks. Glycosyl moieties in the N-terminal region are removed for clarity.

rimonabant (also denoted SR141716A) in a manner nearly identical to wild-type CB1. CB1(T210A)–PGS has, however, a sevenfold lower affinity for the agonist CP55940, consistent with stabilization of an inactive conformation (Extended Data Fig. 2) and in agreement with the original report of the T210A mutation¹⁵. After purifying this construct from Sf9 insect cells (Methods and Extended Data Fig. 3), we obtained LCP microcrystals that diffracted to 2.6 Å resolution, solved the structure by molecular replacement and refined the structure to an R_{free} value of 0.23 (Methods and Extended Data Table 1). In the monoclinic crystals, CB1(T210A)–PGS packs in a manner such that the extracellular-facing ligand-binding region is not involved in lattice contacts (Extended Data Fig. 4a), and the receptor and ligand are well ordered with low overall B factors. Although truncation of the N terminus of CB1 was necessary to form diffraction-quality crystals, such modifications may affect the functional properties of the receptor, as indicated by the variable expression and pharmacology of tissue-specific splice variants in this region¹⁶. Nevertheless, our binding data (Extended Data Fig. 2) and previous precedent¹⁷ show that the basic inverse-agonist- and agonist-binding properties of CB1 are maintained in a receptor lacking the N-terminal 89 residues (which also contains three consensus N-linked glycosylation motifs).

The global structure of the CB1 receptor, with its classical seven-transmembrane fold, is shown in Fig. 1. Using other rhodopsin family (class A) GPCRs as guides¹⁸, the taranabant-bound CB1 structure represents an inactive conformation with respect to G-protein binding, with a canonical ionic lock formed between Arg214^{3,50} and Asp338^{6,30} (distance, 3.4 Å; Ballesteros–Weinstein numbering used in superscript). At the extracellular surface, the second extracellular loop (ECL2) and

the membrane-proximal N-terminal region preceding transmembrane domain 1 (TM1) form a lid over the orthosteric pocket, which almost completely shields taranabant from the solvent (Fig. 1a, b). As is observed in the structure of the lipid-activated GPCR S1P₁ (ref. 19), a gap between TM1 and TM7 in the extracellular leaflet (Fig. 1b) may contribute to a membrane-embedded access channel for lipophilic agonists. Further dilation of the highly conserved residues (Ile119^{1,35}, Phe381^{7,37}, and Met384^{7,40}) that line this channel (Extended Data Fig. 5) would be required to facilitate entry of ligands. Previous molecular dynamics simulations proposed that the endocannabinoid 2-AG enters into the homologous CB2 receptor between TM6 and TM7 (ref. 20); however, these two transmembrane domains are tightly associated in the present structure. Taranabant makes multiple contacts with both TM1 and TM7 and fills the orthosteric pocket directly inside the TM1–TM7 opening, potentially acting as a plug that blocks entry of the endocannabinoid. The extracellular face and lid above the orthosteric pocket contain an abundance of acidic residues, giving a negatively charged surface that will energetically disfavour interaction with negatively charged ligands (Fig. 1c). This feature of CB1 may help to ensure lipid-binding selectivity in a bilayer containing a high concentration of negatively charged phospholipids.

The first part of the N terminus of CB1 observed in the electron density of our crystals begins at E100. The 13 membrane-proximal amino acids that precede TM1 fold over the ligand-binding pocket and interact with TM2, TM3, ECL2, and TM7 (Fig. 2a, b). This region is highly conserved in CB1 (Extended Data Fig. 6) and contributes extensively to interaction with taranabant (as will be discussed). The occluded nature of the CB1 orthosteric pocket was predicted by a study showing that disulfide bond formation between Cys98 and Cys107 modulates orthosteric ligand binding²¹; however, this disulfide bond is either not present or not visible in the current structure (possibly owing to cysteine capping by iodoacetamide). To assess the flexibility of the N-terminal region of CB1, we carried out a 60-ns molecular dynamics simulation of the CB1 structure embedded in an explicit POPC bilayer in the presence and absence of taranabant. In both cases, the N-terminal region was highly stable over the course of the simulation, exhibiting low root mean squared deviation (r.m.s.d.) values comparable to those of the entire transmembrane bundle (Extended Data Fig. 7). These results support the idea that the N-terminal region of CB1 will maintain a conformation similar to the structure observed here, even in the absence of ligand. Other lipid-activated GPCRs that have been structurally characterized (S1P₁ and LPA₁; refs 19 and 22, respectively) contain a disulfide-cross-linked ECL2 structure that is very similar to that of CB1; however, the N-terminal regions of these receptors are markedly different, containing α -helices that sit above the membrane and pack between ECL1 and ECL2 (Fig. 2c, d). The occluded orthosteric pocket of CB1, with the N-terminal region folding over the buried hydrophobic inverse agonist taranabant, is mirrored in the structure of the visual photoreceptor rhodopsin bound to 11-*cis*-retinal²³ (Fig. 2e). A gap between TM1 and TM7 was proposed as part of a channel for uptake and release of the lipophilic 11-*cis*-retinal ligand, based on the structure of the ligand-free opsin in an active conformation²⁴, further paralleling the structure of CB1. The opsin residues Leu40^{1,35}, Ile290^{7,37}, and Phe293^{7,40} surrounding this gap are analogous to CB1 residues Ile119^{1,35}, Phe381^{7,37}, and Met384^{7,40} (Extended Data Fig. 5).

Taranabant is a subtype-selective inverse agonist with an inhibition constant (K_i) of 0.13 nM for CB1 and a K_i of 170 nM for CB2 (ref. 25). Unambiguous electron density at the orthosteric ligand-binding pocket (Extended Data Fig. 4b, c) placed taranabant at an unusual site, towards TM1 and TM7, contrasting with the space occupied by inhibitors of other class A GPCRs, such as the β_2 adrenergic receptor¹⁸ (Fig. 3a). Taranabant adopts a conformation in which the chlorophenyl moiety extends towards TM5, the cyanophenyl buries deeper into the seven transmembrane bundle and the trifluoromethylpyridine projects into the putative access channel between TM1 and TM7 (Fig. 3b).

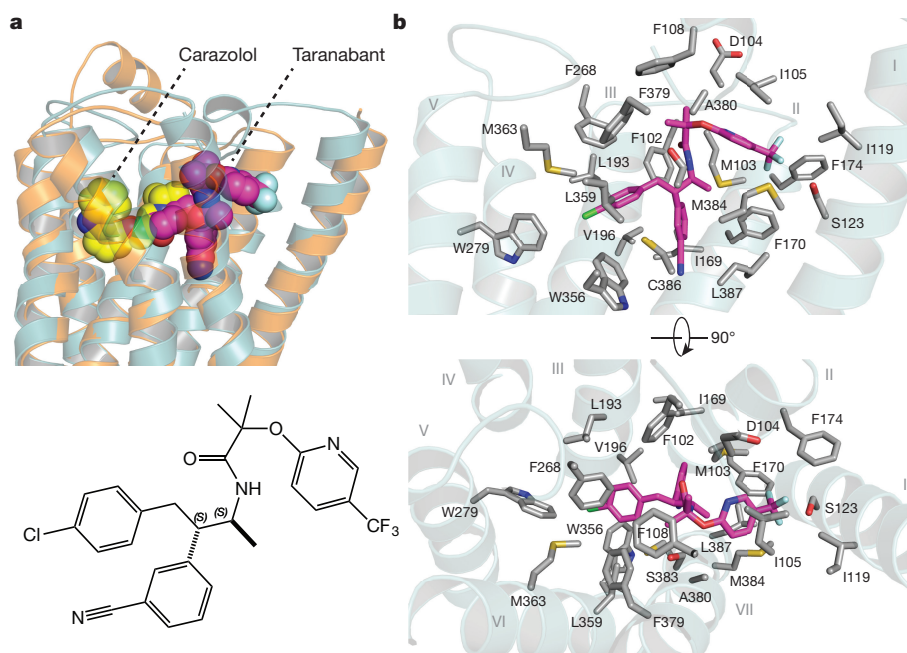


Figure 3 | Binding of taranabant to the CB1 receptor. **a**, Cartoon of CB1 (translucent, teal) with a cartoon of β_2 AR (PDB accession number 2RH1; translucent, orange) superimposed. The r.m.s.d. for the C α positions is 2.6 Å. The beta blocker carazolol is represented by yellow spheres, taranabant by magenta spheres. At the bottom is a 2D representation

of taranabant. **b**, Contact residues with 4 Å of taranabant in the CB1 structure. The receptor side chains are shown as grey sticks, the backbone is a transparent cartoon and taranabant is in magenta. Top view is from within the plane of the bilayer (TM6 and TM7 cartoons removed for clarity), bottom view is from the extracellular space.

The orthosteric binding pocket of CB1 is highly hydrophobic, as is expected for a lipid-activated receptor. Of the 24 residues within 4 Å of the ligand, there are only three polar side chains: Asp104, whose acidic side chain points towards the extracellular space; Ser123^{1,39} near the access channel that forms a polar contact with the trifluoromethyl group of taranabant; and Ser383^{7,39}, which has been implicated in agonist binding²⁶. By contrast, a large number of hydrophobic residues (including six Phe, three Met, two Trp, three Leu, and three Ile side chains) line the orthosteric pocket and make a variety of hydrophobic contacts with taranabant, burying 1,109 Å² of surface area (Fig. 3b). All of the taranabant contact residues on CB1 are absolutely conserved across the vertebrate lineage, with the exception of Ile105, which can be replaced by Met (Extended Data Fig. 8). The major divergence between CB1 and CB2 within the subset of binding residues lies in the membrane-proximal N-terminal region, where Phe102, Met103, Asp104, Ile105, and Phe108 make van der Waals contacts with taranabant. The subtype selectivity of taranabant for CB1 may arise from the divergence of this region between CB1 and CB2.

Taranabant (Fig. 3a) and rimonabant (Fig. 4a) have related chemical structures and similar conformational properties in isolation²⁷. Docking of rimonabant with the CB1 crystal structure yielded a low-energy pose that overlaps almost completely with that of taranabant, contacting the same constellation of residues (Fig. 4a). This supports the use of the current structure to analyse the binding modes of both ligands. Mutagenesis studies have identified several residues whose mutation caused a loss in taranabant and/or rimonabant binding affinity^{27–30}. Indeed, many of these residues are in contact with the ligand in the CB1 structure, including Phe170^{2,57}, Phe174^{2,61}, Leu193^{3,29}, Trp279^{5,43}, Trp356^{6,48}, Phe379^{7,35}, and Leu387^{7,42}. However, several residues on TM3 and TM5 (for example, Lys192^{3,28}, Phe200^{3,36}, and Tyr275^{3,39}) are not within contact distance of taranabant and appear to make indirect contributions to binding, through structural stabilization or influence of the conformational equilibrium of CB1.

To gain insight into the initial recognition of agonists by the CB1 receptor, we docked THC (a partial agonist⁵) into our crystal structure coordinates using the program Glide (see Methods). The top docking

poses have the tricyclic core of THC binding between TM1, TM2 and TM7 (as with taranabant), with the C3 alkyl chain overlapping with the chlorophenyl moiety of taranabant and extending towards Trp356^{6,48} (Fig. 4b). Conformational changes in this residue and its surroundings have been proposed as a trigger for CB1 activation, and mutation to alanine leads to enhanced stimulation (E_{\max}) by CB1 agonists³¹. Previous mutagenesis experiments have also identified Phe174^{2,61}, Leu193^{3,29}, and Ser383^{7,39} as important residues for binding of THC or related agonists such as CP55940 (refs 26, 32). These residues are either in contact with or in close proximity to the preferred docking pose of THC. One caveat to these calculations is that the inactive structure of CB1 is not ideal for predicting high-affinity agonist interactions. It should, however, be noted that the crystallization construct (stabilized in an inactive conformation) still displays significant affinity for CP55940 (Extended Data Fig. 2). Finally, Cys355^{6,47} on the bilayer-facing side of TM6 was reported to form a covalent adduct with a THC analogue that possesses a reactive group at the end of the C3-pentyl chain³³. Starting with our THC pose, such cross-linking would require rotation of TM6 at the orthosteric pocket during CB1 activation and consequent disruption of the packing around Trp356^{6,48}.

While our manuscript was under review, a crystal structure of human CB1 was reported³⁴ bound to the antagonist AM6538, which closely resembles rimonabant but has a nitrate group substituted on 'arm 2' of the rimonabant core (that is, the chlorophenyl moiety in Fig. 4a). Although the taranabant-bound crystal structure reported here and the AM6538-bound structure are in general agreement (Extended Data Fig. 9a), there are several differences that may be important for functional interpretation and prediction. Notably, the electron density for the ligand and the important N-terminal region is weak in the AM6538-bound structure, with high B factors in the refined model (average $B = 134.3$ Å² for residues 99–112 and $B = 119.5$ Å² for the ligand). By contrast, the equivalent region in our taranabant-bound structure is very well ordered, with good density and much lower B factors (average $B = 61.7$ Å² for residues 100–112 and $B = 42.0$ Å² for the ligand) (Extended Data Fig. 9b, c). The lack of clear density and resulting model ambiguity for the N-terminal region in the AM6538-bound structure

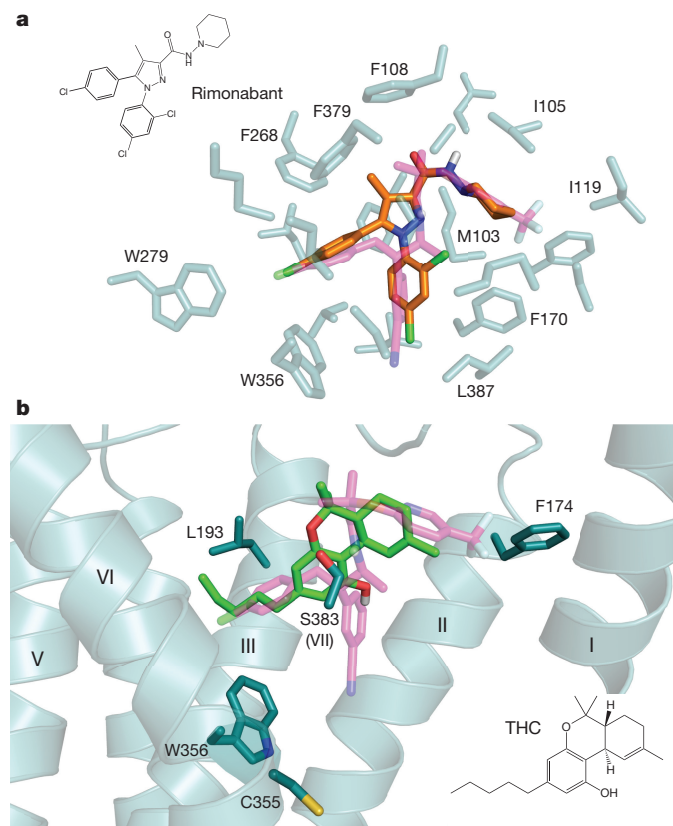


Figure 4 | Docking of rimonabant and THC to the CB1 receptor.

a, Overlay of the crystal structure pose of taranabant (transparent magenta sticks) with the top-scoring docking pose of rimonabant shown using orange sticks (see Methods). The contact residues within 4 Å of taranabant are shown as transparent teal sticks. The 2D structure of rimonabant is shown at upper left. **b**, Top-scoring docking pose of THC is shown as light green sticks, along with taranabant (transparent magenta sticks). Selected residues important for the binding of THC and agonist activity are shown as teal sticks. TM7 cartoon is removed for clarity. The 2D structure of THC is shown on the bottom right.

may limit its utility for predicting the binding modes of other ligands. This is apparent in the erroneous docking prediction for taranabant, in which arm 1 and arm 2 (chlorophenyl and cyanophenyl groups) are swapped relative to their experimentally determined binding positions reported herein. Further biochemical and computational studies will be required to establish the relative utility of these two crystal structures as templates for ligand docking and design.

GPCRs adopt multiple conformations, creating a complex energy landscape that allows the binding of different ligands to modulate different intracellular effectors, such as G proteins and arrestin³⁵. CB1 has considerable agonist-independent constitutive activity³⁵ and exhibits subtle and paradoxical pharmacological properties; it is antagonized by cannabidiol (a molecule that, but for a bond disconnection, is near-identical to THC)⁵ and inhibited by the compound ORG27569, which allosterically increases agonist affinity but decreases subsequent receptor activation³⁷. Understanding these phenomena will require additional structures of CB1 in different conformational states, bound to a range of ligands (both orthosteric and allosteric) of differing efficacy. Our structure of CB1 bound to taranabant represents a step in this direction and provides a crystallographic basis for computational design of cannabinoid system modulators.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 October; accepted 8 November 2016.

Published online 16 November 2016.

- Mechoulam, R. & Parker, L. A. The endocannabinoid system and the brain. *Annu. Rev. Psychol.* **64**, 21–47 (2013).
- Lynch, M. E. & Ware, M. A. Cannabinoids for the treatment of chronic non-cancer pain: an updated systematic review of randomized controlled trials. *J. Neuroimmune Pharmacol.* **10**, 293–301 (2015).
- Reddy, D. S. & Golub, V. M. The pharmacological basis of cannabis therapy for epilepsy. *J. Pharmacol. Exp. Ther.* **357**, 45–55 (2016).
- Kim, J., Li, Y. & Watkins, B. A. Endocannabinoid signaling and energy metabolism: a target for dietary intervention. *Nutrition* **27**, 624–632 (2011).
- Pertwee, R. G. The diverse CB1 and CB2 receptor pharmacology of three plant cannabinoids: Δ^9 -tetrahydrocannabinol, cannabidiol and Δ^9 -tetrahydrocannabivarin. *Br. J. Pharmacol.* **153**, 199–215 (2008).
- Vemuri, V. K. & Makriyannis, A. Medicinal chemistry of cannabinoids. *Clin. Pharmacol. Ther.* **97**, 553–558 (2015).
- Howlett, A. C. et al. Cannabinoid physiology and pharmacology: 30 years of progress. *Neuropharmacology* **47** (Suppl. 1), 345–358 (2004).
- Wilson, R. I. & Nicoll, R. A. Endocannabinoid signaling in the brain. *Science* **296**, 678–682 (2002).
- Fowler, C. J. Transport of endocannabinoids across the plasma membrane and within the cell. *FEBS J.* **280**, 1895–1904 (2013).
- Tam, J. et al. Peripheral cannabinoid-1 receptor inverse agonism reduces obesity by reversing leptin resistance. *Cell Metab.* **16**, 167–179 (2012).
- Jourdan, T. et al. Activation of the Nlrp3 inflammasome in infiltrating macrophages by endocannabinoids mediates beta cell loss in type 2 diabetes. *Nat. Med.* **19**, 1132–1140 (2013).
- Gaoni, Y. & Mechoulam, R. Isolation, structure, and partial synthesis of an active constituent of hashish. *J. Am. Chem. Soc.* **86**, 1646–1647 (1964).
- Janero, D. R. & Makriyannis, A. Cannabinoid receptor antagonists: pharmacological opportunities, clinical experience, and translational prognosis. *Expert Opin. Emerg. Drugs* **14**, 43–65 (2009).
- Yin, J. et al. Structure and ligand-binding mechanism of the human OX1 and OX2 orexin receptors. *Nat. Struct. Mol. Biol.* **23**, 293–299 (2016).
- D'Antona, A. M., Ahn, K. H. & Kendall, D. A. Mutations of CB1 T210 produce active and inactive receptor forms: correlations with ligand affinity, receptor stability, and cellular localization. *Biochemistry* **45**, 5606–5617 (2006).
- González-Mariscal, I. et al. Human CB1 receptor isoforms, present in hepatocytes and β -cells, are involved in regulating metabolism. *Sci. Rep.* **6**, 33302 (2016).
- Andersson, H., D'Antona, A. M., Kendall, D. A., Von Heijne, G. & Chin, C. N. Membrane assembly of the cannabinoid receptor 1: impact of a long N-terminal tail. *Mol. Pharmacol.* **64**, 570–577 (2003).
- Rosenbaum, D. M. et al. GPCR engineering yields high-resolution structural insights into β 2-adrenergic receptor function. *Science* **318**, 1266–1273 (2007).
- Hanson, M. A. et al. Crystal structure of a lipid G protein-coupled receptor. *Science* **335**, 851–855 (2012).
- Hurst, D. P. et al. A lipid pathway for ligand binding is necessary for a cannabinoid G protein-coupled receptor. *J. Biol. Chem.* **285**, 17954–17964 (2010).
- Fay, J. F. & Farrens, D. L. The membrane proximal region of the cannabinoid receptor CB1 N-terminus can allosterically modulate ligand affinity. *Biochemistry* **52**, 8286–8294 (2013).
- Chrencik, J. E. et al. Crystal structure of antagonist bound human lysophosphatidic acid receptor 1. *Cell* **161**, 1633–1643 (2015).
- Palczewski, K. et al. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739–745 (2000).
- Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H.-W. & Ernst, O. P. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **454**, 183–187 (2008).
- Fong, T. M. et al. Antiobesity efficacy of a novel cannabinoid-1 receptor inverse agonist, *N*-[(1*S*,2*S*)-3-(4-chlorophenyl)-2-(3-cyanophenyl)-1-methylpropyl]-2-methyl-2-[[5-(trifluoromethyl)pyridin-2-yl]oxy]propanamide (MK-0364), in rodents. *J. Pharmacol. Exp. Ther.* **321**, 1013–1022 (2007).
- Kapur, A. et al. Mutation studies of Ser7.39 and Ser2.60 in the human CB1 cannabinoid receptor: evidence for a serine-induced bend in CB1 transmembrane helix 7. *Mol. Pharmacol.* **71**, 1512–1524 (2007).
- Lin, L. S. et al. Conformational analysis and receptor docking of *N*-[(1*S*,2*S*)-3-(4-chlorophenyl)-2-(3-cyanophenyl)-1-methylpropyl]-2-methyl-2-[[5-(trifluoromethyl)pyridin-2-yl]oxy]propanamide (taranabant, MK-0364), a novel, acyclic cannabinoid-1 receptor inverse agonist. *J. Med. Chem.* **51**, 2108–2114 (2008).
- Shim, J.-Y., Bertalovitz, A. C. & Kendall, D. A. Probing the interaction of SR141716A with the CB1 receptor. *J. Biol. Chem.* **287**, 38741–38754 (2012).
- Sitkoff, D. F. et al. Cannabinoid CB(1) receptor ligand binding and function examined through mutagenesis studies of F200 and S383. *Eur. J. Pharmacol.* **651**, 9–17 (2011).

30. Hurst, D. P. *et al.* *N*-(piperidin-1-yl)-5-(4-chlorophenyl)-1-(2,4-dichlorophenyl)-4-methyl-1H-pyrazole-3-carboxamide (SR141716A) interaction with LYS 3.28(192) is crucial for its inverse agonism at the cannabinoid CB1 receptor. *Mol. Pharmacol.* **62**, 1274–1287 (2002).
31. McAllister, S. D. *et al.* Structural mimicry in class A G protein-coupled receptor rotamer toggle switches: the importance of the F3.36(201)/W6.48(357) interaction in cannabinoid CB1 receptor activation. *J. Biol. Chem.* **279**, 48024–48037 (2004).
32. Shim, J.-Y., Bertalovitz, A. C. & Kendall, D. A. Identification of essential cannabinoid-binding domains: structural insights into early dynamic events in receptor activation. *J. Biol. Chem.* **286**, 33422–33435 (2011).
33. Picone, R. P. *et al.* (–)-7'-Isothiocyanato-11-hydroxy-1,1'-dimethylheptylhexahydrocannabinol (AM841), a high-affinity electrophilic ligand, interacts covalently with a cysteine in helix six and activates the CB1 cannabinoid receptor. *Mol. Pharmacol.* **68**, 1623–1635 (2005).
34. Hua, T. *et al.* Crystal structure of the human cannabinoid receptor CB1. *Cell* **167**, 750–762.e14 (2016).
35. Manglik, A. *et al.* Structural insights into the dynamic process of β 2-adrenergic receptor signaling. *Cell* **161**, 1101–1111 (2015).
36. Console-Bram, L., Marcu, J. & Abood, M. E. Cannabinoid receptors: nomenclature and pharmacological principles. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **38**, 4–15 (2012).
37. Price, M. R. *et al.* Allosteric modulation of the cannabinoid CB1 receptor. *Mol. Pharmacol.* **68**, 1484–1495 (2005).

Acknowledgements We thank the staff of the GM/CA-CAT beamline 23ID at the Advanced Photon Source (APS) for support during data collection. This project was supported by Welch Foundation grant (I-1770 to D.M.R.) and a Packard Foundation Fellowship (D.M.R.). APS is a US Department of Energy Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory (DE-AC02-06CH11357).

Author Contributions Z.S. developed the CB1 construct and purification; expressed, purified and crystallized the receptor; collected diffraction data; and solved and refined the structure. J.Y. assisted with crystallographic refinement. K.C. performed ligand binding assays on CB1 constructs. M.G. carried out computational docking calculations. L.C. assisted with construct design and purification. J.W. performed and supervised computational docking calculations and molecular dynamics simulations. D.M.R. supervised the overall project, assisted with collection of diffraction data, and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.M.R. (dan.rosenbaum@utsouthwestern.edu).

Reviewer Information *Nature* thanks A. Christopoulos, G. Kunos and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Cloning, expression and purification. The wild-type human CB1 receptor gene (Uniprot Entry: P21554) was cloned into a modified pFastBac (Invitrogen) baculovirus expression vector with the haemagglutinin (HA) signal sequence followed by a Flag epitope tag at the N terminus and a 10× His-tag at the C terminus. To facilitate receptor crystallization, the 76N-terminal residues were removed, a TEV protease recognition site was introduced before residue Lys90, and the 51 C-terminal residues were deleted (truncation after Pro421). Residues 302–332 in the CB1 intracellular loop 3 (ICL3) were replaced with a synthetic DNA fragment containing the 196-amino acid coding sequence of *P. abyssi* glycogen synthase (PDB accession number: 2FBW). Finally, the mutation T210A was introduced by an adapted Multi-site Quickchange protocol (Stratagene).

The final CB1 (T210A)–PGS fusion construct was transfected into DH10Bac to produce a recombinant baculovirus with the Bac-to-Bac system (Invitrogen). The recombinant baculovirus was used to infect Sf9 insect cell culture at a cell density of 2.5×10^6 cells per ml^{-1} , with 1 μM taranabant (Tocris) added to the medium. Infected cells were grown for 60 h at 27 °C before harvesting, and the cell pellets were stored at -80°C for future use.

Sf9 cell membranes were disrupted by thawing frozen cell pellets in a hypotonic buffer containing 10 mM Tris pH 7.5, 1 mM EDTA, 160 $\mu\text{g ml}^{-1}$ benzamidine, 100 $\mu\text{g ml}^{-1}$ leupeptin, 2 mg ml^{-1} iodoacetamide and 1 μM taranabant. The cell membranes were centrifuged at 10,000 g for 20 min at 4 °C. Membrane pellets were solubilized in a buffer containing 50 mM HEPES, pH 7.5, 500 mM NaCl, 1% (w/v) *n*-dodecyl β -D-maltopyranoside (DDM; Anatrace), 0.2% sodium cholate, 0.2% cholesteryl hemi-succinate (CHS), 10% glycerol, 160 $\mu\text{g ml}^{-1}$ benzamidine, 100 $\mu\text{g ml}^{-1}$ leupeptin, 2 mg ml^{-1} iodoacetamide and 10 μM taranabant for 1 h at 4 °C. The supernatant was isolated after ultra-centrifugation for 30 min at 100,000 g and incubated with Ni-NTA agarose beads (GE Healthcare) in batch for 3 h at 4 °C. After binding, the beads were collected by centrifugation at 100g and washed with five volumes of Ni-NTA wash buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 0.05% (w/v) DDM, 0.01% sodium cholate, 0.01% CHS, 10% glycerol, 50 mM imidazole, 160 $\mu\text{g ml}^{-1}$ benzamidine, 100 $\mu\text{g ml}^{-1}$ leupeptin and 1 μM taranabant). After transfer to a gravity column, beads were washed with 15 column volumes of Ni-NTA wash buffer, and receptor protein was eluted in Ni-NTA wash buffer with 200 mM imidazole and 2 mM calcium. The eluted protein was then loaded by gravity flow over anti-Flag M1 affinity resin. Detergent was exchanged from 0.05% DDM to 0.05% lauryl maltose neopentyl glycol (LMNG) on the M1 resin. Finally the pure receptor was eluted with 0.2 mg ml^{-1} Flag peptide and 5 mM EDTA. TEV protease (1:10 w/w) and PNGase F were added to the eluate, and protein was incubated at 4 °C overnight. Finally, the receptor was run on a Superdex 200 size-exclusion column (GE Healthcare) with buffer containing 20 mM HEPES, pH 7.5, 150 mM NaCl, 0.05% LMNG, and 1 μM taranabant.

Differential scanning fluorimetry. Protein samples were purified and prepared in the absence of ligand (apo), with taranabant, or with rimonabant, as described above. Differential scanning fluorimetry assays were performed in 96-well PCR plates using a real-time PCR machine (CFX96, Bio-Rad). Standard assay conditions (25 μl) were 25 mM HEPES pH 7.5, 150 mM NaCl, 0.01% LMNG, 0.002% CHS and 10 μM of the corresponding ligands. The protein concentration was 2 μM and the BODIPY FL-L-cystine dye³⁸ was added at 2 μM final concentration. All reactions were incubated at 4 °C for 20 min before scanning in the PCR machine. The fluorescence was measured at 0.5 °C temperature intervals from 4 °C to 90 °C by using the FAM filter set (450–490 nm excitation, 515–530 nm emission).

Crystallization. Purified receptor was concentrated to 55 mg ml^{-1} using a 100-kDa cut-off Vivaspin column (Sartorius), and crystallized using the LCP method. The concentrated receptor was reconstituted into a lipid mixture containing monoolein plus 10% (w/w) cholesterol (Sigma), at a ratio of 2:3 receptor to lipid (by weight). Mixing was performed at room temperature using a syringe mixing apparatus as previously described³⁹. The mesophase was dispensed in 40-nl drops onto 96-well glass plates and overlaid with 800 nl precipitant solution using a Gryphon LCP robot (Art Robbins Instruments). Crystals grew to full size after 2 weeks at 20 °C in the following overlay precipitant condition: 31% PEG 400, 100 mM sodium citrate pH 5.5, 100 mM magnesium sulfate. The crystals were harvested from LCP setups using MiTeGen loops and cryoprotected in liquid nitrogen.

Data collection and processing. X-ray diffraction data were collected at GM/CA-CAT beamline 23ID-B at the Advanced Photon Source (APS), Argonne National Laboratory, equipped with an Eiger 16M detector. Datasets were acquired using a beam size of 20 μm with 1.033-Å wavelength X-rays. For each crystal, fifty 0.4° oscillation images were collected, with 1-s exposure and without attenuation of the beam. Owing to radiation damage of crystals, a 97% complete diffraction data was merged from 42 crystals and scaled using HKL3000⁴⁰. The dataset was

processed in space group $P2_1$, and a resolution cut-off of 2.6 Å was selected by examining $CC_{1/2}$ values after anisotropy correction in HKL3000.

Structure determination and refinement. The structure of CB1 (T210A)–PGS in complex with taranabant was solved by molecular replacement with Phaser⁴¹ using human S1P₁ receptor¹⁹ (PDB: 3V2Y) and PGS⁴² (PDB: 2BFW) as independent search models. The solution was improved through iterations of manual building in Coot⁴³, followed by refinement using Refmac5 (ref. 44). Translation-libration-screw refinement was used to model atomic displacement factors. Refinement parameters for the taranabant ligand were generated using the PRODRG⁴⁵ web server. The resulting statistics for data collection and refinement are included in Extended Data Table 1. The final structure had 96.6% of residues in the favoured region of the Ramachandran plot, 3.4% in the allowed region, and 0 residues disallowed. Figures were prepared using Pymol (Schrodinger LLC). The electrostatic potential surface shown in Fig. 1c was calculated using APBS⁴⁶.

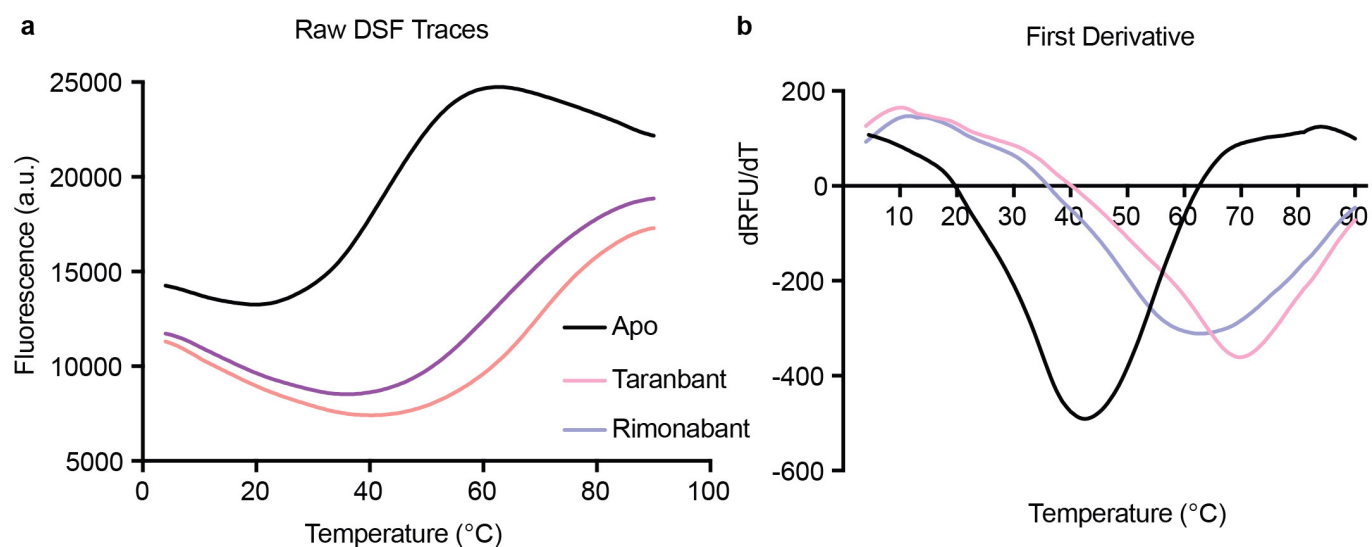
Binding of ligands to the CB1 receptor. Ligand-binding experiments on membranes containing CB1 wild-type, CB1–PGS, and CB1 (T210A)–PGS were carried out based on a previously published protocol²⁸. Sf9 cells expressing each construct (without any ligand present) were used to generate membranes by Dounce homogenization and differential centrifugation¹⁸. Saturation binding was carried out by incubating 1.5–5 μg of membranes with different concentrations of [³H]SR141716A (54 Ci mmol^{-1} ; Perkin-Elmer) between 0.05 and 25.6 nM in assay buffer (25 mM Tris pH 7.5, 5 mM MgCl₂, 1 mM EDTA) containing 0.1% protease-free BSA in a final volume of 250 μl per tube. Reactions were incubated at 30 °C for 1 h and then quenched with 250 μl assay buffer with 5% BSA. Non-specific binding was determined using reactions containing 1 μM unlabelled ligand. Reactions were separated on a vacuum manifold using GF/C filters (pre-soaked in assay buffer supplemented with 0.5% polyethylenimine) to retain membranes and discard unbound ligand. After washing four times with cold assay buffer, bound radioactivity was quantified using a scintillation counter. For competition-binding experiments, aliquots of membranes were incubated with 3 nM [³H]SR141716A, and varying concentrations of competitor ligands (taranabant or CP55940) were included in the binding reactions. All binding experiments were carried out as three independent experiments, each performed in duplicate. Data analysis and fitting was performed with GraphPad Prism (GraphPad Software Inc.).

Molecular dynamics simulations. The system used for molecular dynamics simulation consisted of one copy of CB1 receptor (PGS domain removed), taranabant, 240 POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) molecules, 48 Na⁺, 57 Cl[−], and 17,087 water molecules. Molecular dynamics simulations were performed with periodic boundary condition to produce isothermal-isobaric ensembles using the modified PMEMD.CUDA program in AMBER 14 (ref. 47). Temperature was regulated using Langevin dynamics⁴⁸ with a collision frequency of 5 ps^{−1}. Pressure was regulated using the isotropic position scaling algorithm with the pressure relaxation time set to 1.0 ps. The integration of the equations of motion was conducted at a time step of 1 fs for the relaxation phase and 2 fs for the equilibrium and sampling phases. After a 5-ns equilibration, a 55-ns molecular dynamics simulation was performed at 298 K, 1 bar to produce constant temperature and pressure ensembles. The transmembrane helices were very stable in both simulations and the mean r.m.s.d. values were 1.52 ± 0.13 and 1.45 ± 0.23 Å for the apo and complex forms, respectively. The r.m.s.d. values of the membrane-proximal N-terminal region of the complex form (0.96 ± 0.24 Å) were smaller than those of the apo form (1.28 ± 0.19 Å).

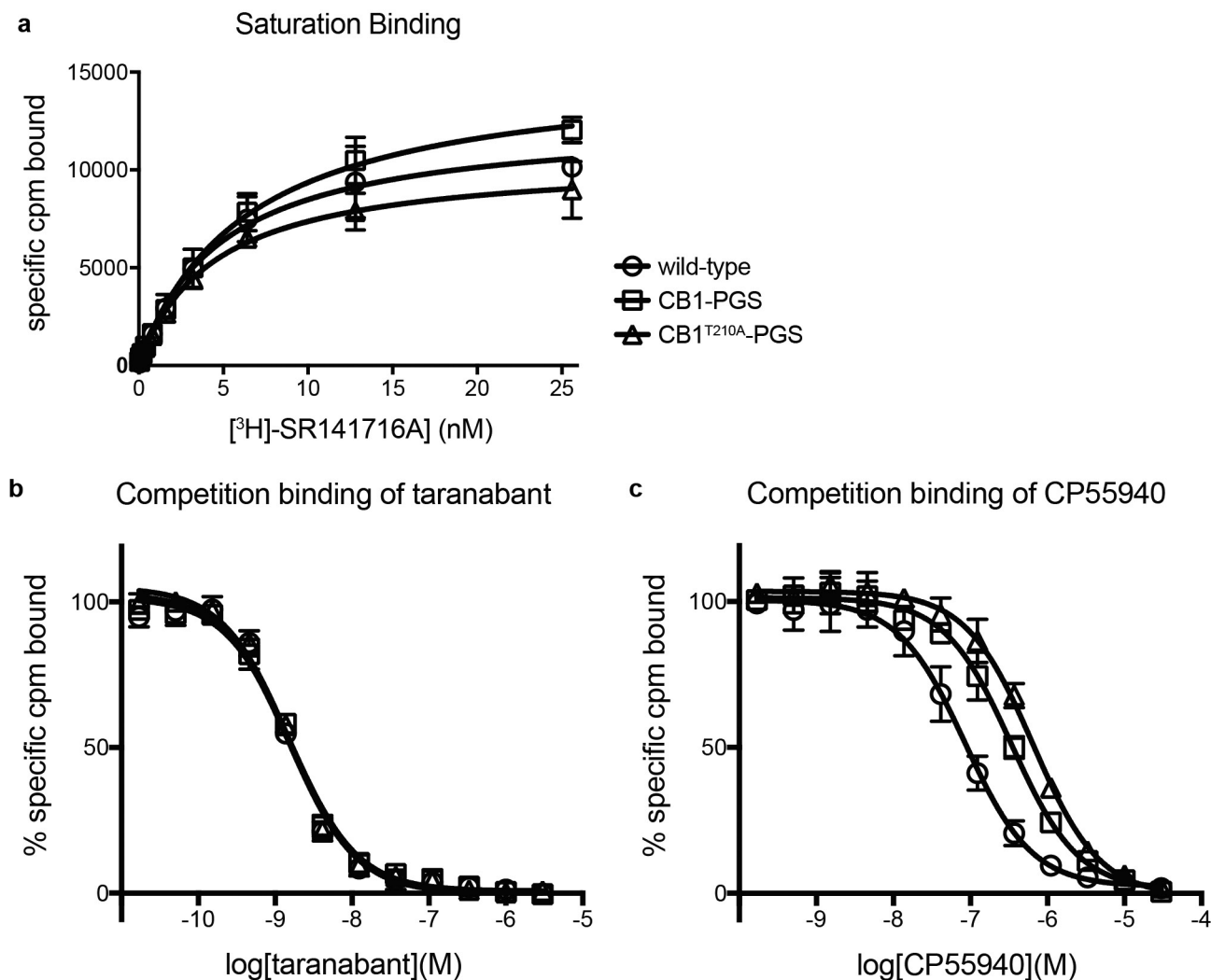
Docking of rimonabant and THC. Molecular docking was performed for taranabant, rimonabant, and THC using Glide^{49,50}, implemented in the Schrodinger software package (<http://www.schrodinger.com>). Different protocols of receptor preparation, grid generation and flexible ligand docking were evaluated and the one that produced the best docking scores was adopted. The optimal Glide protocol for CB1 included: only optimize hydrogen atoms in the receptor preparation; allow hydroxyl and thiol groups of Thr197, Ser383 and Cys386 to be rotatable; use the standard precision scoring function. We first tested our docking protocol by re-docking the taranabant ligand from the crystal structure. The best docking scores were -12.76 and -12.59 kcal mol^{-1} for the crystal conformation and a 3D conformation generated without any initial bias using the Concord program (<http://www.certara.com>), respectively. The r.m.s.d. between the crystal structure and docking pose was 0.55 Å for the Concord conformation. Next, the antagonist rimonabant and the partial agonist THC were docked to the binding pocket using the same protocol. The docking scores of the best docking poses were -8.99 and -9.36 kcal mol^{-1} for rimonabant and THC, respectively.

Data availability. Atomic coordinates and structure factors for the reported crystal structure have been deposited in the Protein Data Bank (PDB) under the accession code 5U09. All other data are available from the corresponding author upon reasonable request.

38. Hofmann, L., Gulati, S., Sears, A., Stewart, P. L. & Palczewski, K. An effective thiol-reactive probe for differential scanning fluorimetry with a standard real-time polymerase chain reaction device. *Anal. Biochem.* **499**, 63–65 (2016).
39. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nat. Protocols* **4**, 706–731 (2009).
40. Otwinowski, Z. & Minor, W. Processing of X-ray data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
41. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
42. Horcajada, C., Guinovart, J. J., Fita, I. & Ferrer, J. C. Crystal structure of an archaeal glycogen synthase: insights into oligomerization and substrate binding of eukaryotic glycogen synthases. *J. Biol. Chem.* **281**, 2923–2931 (2006).
43. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
44. Skubák, P., Murshudov, G. N. & Pannu, N. S. Direct incorporation of experimental phase information in model refinement. *Acta Crystallogr. D* **60**, 2196–2201 (2004).
45. Schüttelkopf, A. W. & van Aalten, D. M. F. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D* **60**, 1355–1363 (2004).
46. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
47. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
48. Izaguirre, J. A., Catarella, D. P., Wozniak, J. M. & Skeel, R. D. Langevin stabilization of molecular dynamics. *J. Chem. Phys.* **114**, 2090–2098 (2001).
49. Friesner, R. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47**, 1739–1749 (2004).
50. Friesner, R. A. *et al.* Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **49**, 6177–6196 (2006).



Extended Data Figure 1 | Differential scanning fluorimetry on purified CB1-PGS. **a**, Raw differential scanning fluorimetry traces of the receptor in the apo state or bound to each antagonist. **b**, First derivative analysis of data in **a**.

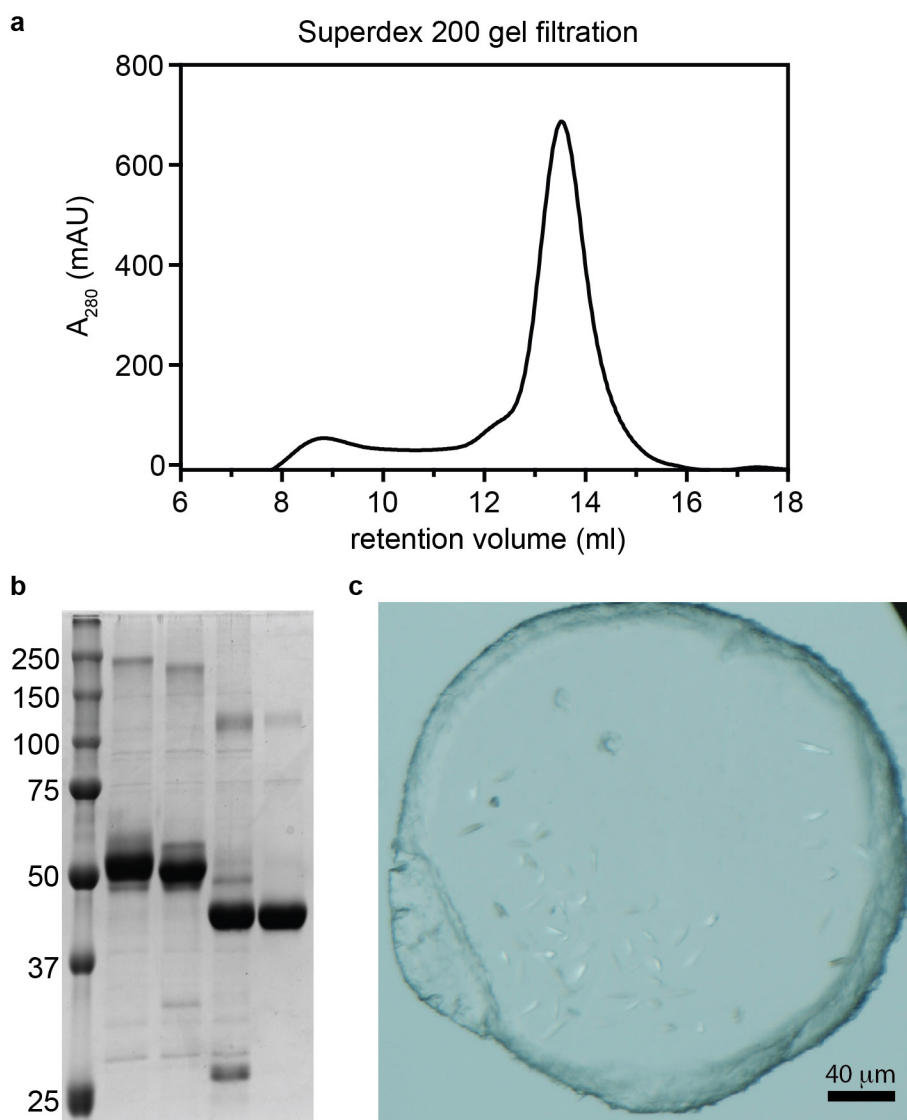


Extended Data Figure 2 | Ligand-binding properties of CB1 constructs.

a, Saturation binding of the antagonist [^3H]SR141716A (tritiated rimonabant radioligand) to wild-type CB1, CB1-PGS, and CB1(T210A)-PGS. Error bars represent s.d. for three separate experiments, each performed in duplicate. The fitted K_d values (\pm s.e.m.) for these three constructs are 4.8 ± 0.7 nM, 6.3 ± 0.6 nM, and 4.4 ± 0.5 nM, respectively.

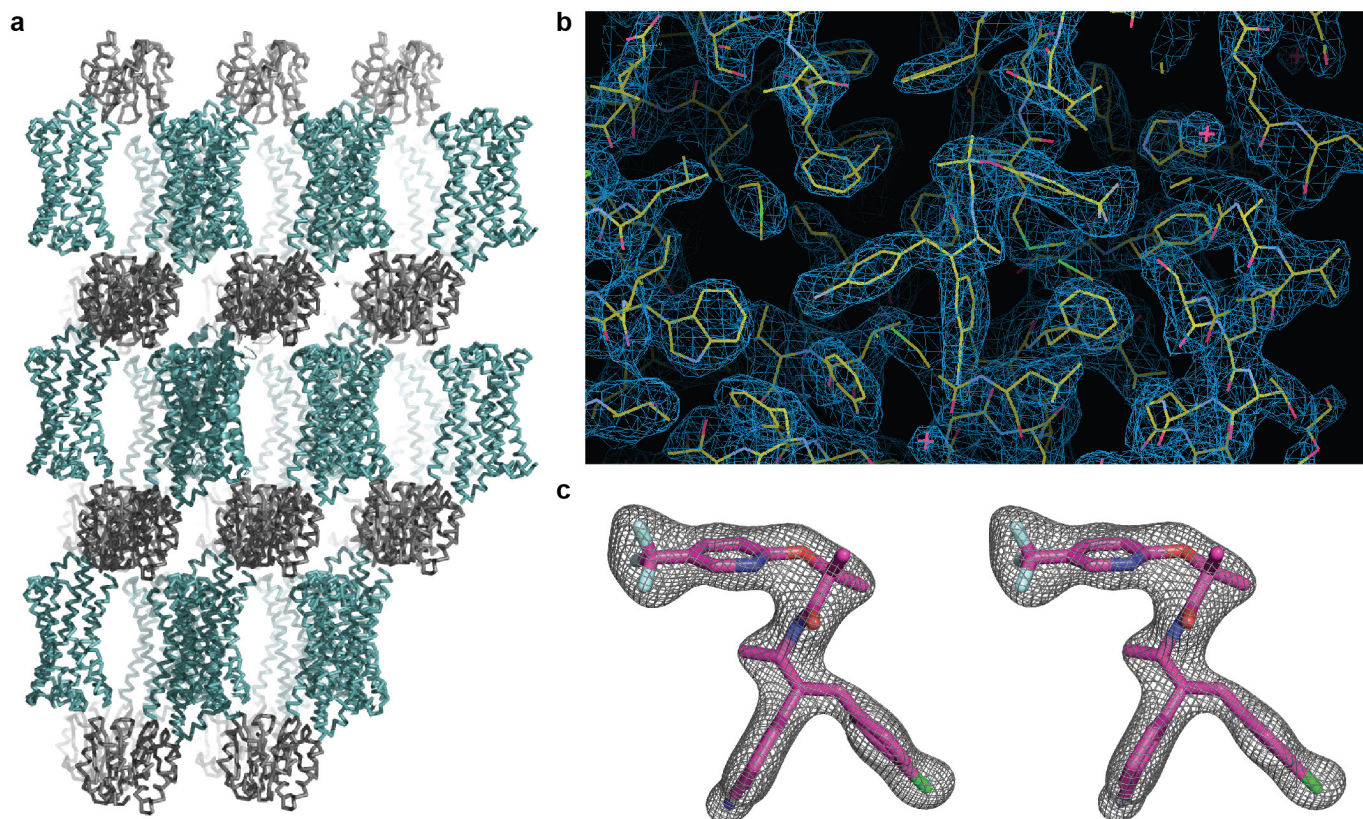
b, Competition binding of taranabant to the wild-type CB1 receptor, CB1-PGS, and CB1(T210A)-PGS. Error bars represent s.d. for three

separate experiments, each performed in duplicate. The K_i values (\pm s.e.m.) of the three constructs for taranabant are 0.94 ± 0.17 nM, 1.10 ± 0.16 nM, and 0.91 ± 0.16 nM, respectively. **c**, Competition binding of the agonist CP55940 to the wild-type CB1 receptor, CB1-PGS, and CB1(T210A)-PGS. Error bars represent s.d. for three separate experiments, each performed in duplicate. The K_i values (\pm s.e.m.) of the three constructs for CP55940 are 53 ± 12 nM, 230 ± 43 nM, and 384 ± 62 nM, respectively.



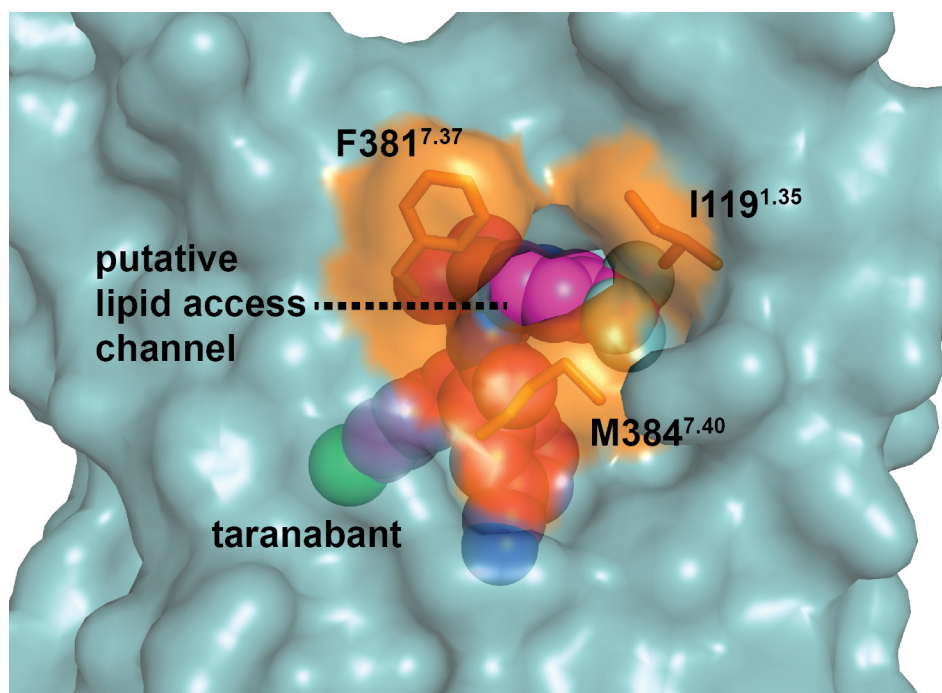
Extended Data Figure 3 | Purification and crystallization of CB1(T210A)–PGS. **a**, Superdex 200 gel-filtration trace of receptor after Ni immobilized metal-affinity chromatography (IMAC) and M1 anti-Flag chromatography (see Methods). **b**, SDS–PAGE analysis of samples at different stages of purification. The five lanes from left to right are:

markers (molecular mass in kDa at left); IMAC/Flag-purified receptor; same sample after treatment with PNGaseF; receptor after TEV protease cleavage (removing 89 N-terminal amino acids); final sample after Superdex 200 gel filtration. **c**, Light microscopy image showing examples of LCP microcrystals of CB1(T210A)–PGS used to collect diffraction data.

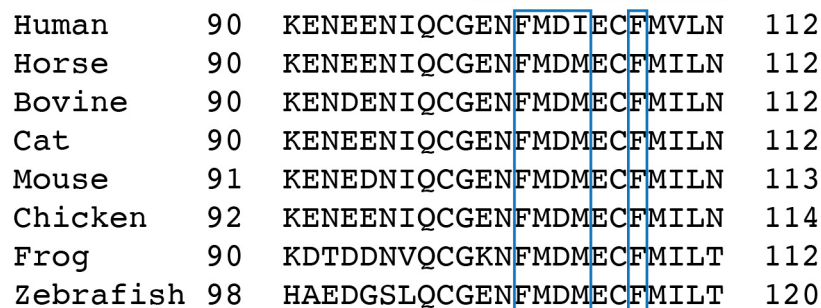


Extended Data Figure 4 | Packing and electron density in the CB1(T210A)-PGS crystals. **a**, Lattice packing interactions in the monoclinic crystals of CB1(T210A)-PGS. Protomers are shown as ribbons, with the receptor component of the fusion protein coloured teal and the PGS domain coloured grey. **b**, $2F_o - F_c$ electron density map

(contoured at 1.2σ) of taranabant and the surrounding ligand-binding residues. Protein and ligand are represented as yellow sticks. **c**, Stereo view of $2F_o - F_c$ electron density (contoured at 1.5σ) for only the ligand taranabant (magenta sticks).



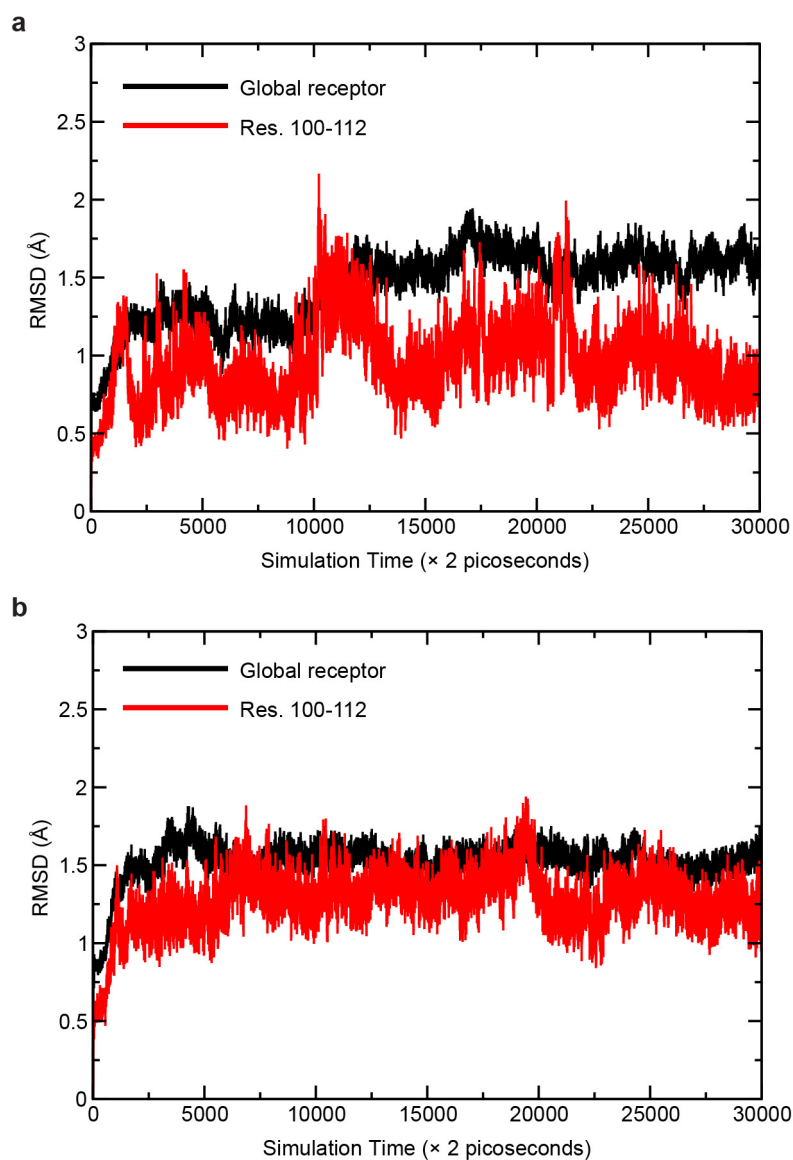
Extended Data Figure 5 | Residues lining the putative lipid access channel of CB1. The receptor is shown as a teal transparent surface, and taranabant is in magenta spheres. The three residues lining the channel are shown as orange sticks and their solvent-accessible surfaces are coloured orange.



Human	90	KENEENIQCGENFMDIECFMVLN	112
Horse	90	KENEENIQCGENFMDMECFMILN	112
Bovine	90	KENDENIQCGENFMDMECFMILN	112
Cat	90	KENEENIQCGENFMDMECFMILN	112
Mouse	91	KENEDNIQCGENFMDMECFMILN	113
Chicken	92	KENEENIQCGENFMDMECFMILN	114
Frog	90	KDTDDNVQCGKNFMDMECFMILT	112
Zebrafish	98	HAEDGSLQCGENFMDMECFMILT	120

Extended Data Figure 6 | Sequence alignment of the membrane-proximal N-terminal region of CB1 from different vertebrate species. 'Frog' is *Xenopus laevis*. The red bar (top) indicates the part of this region that is structured and visible in the electron density of the CB1 crystals.

The blue box denotes positions that make contact with taranabant. Alignment was performed using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).

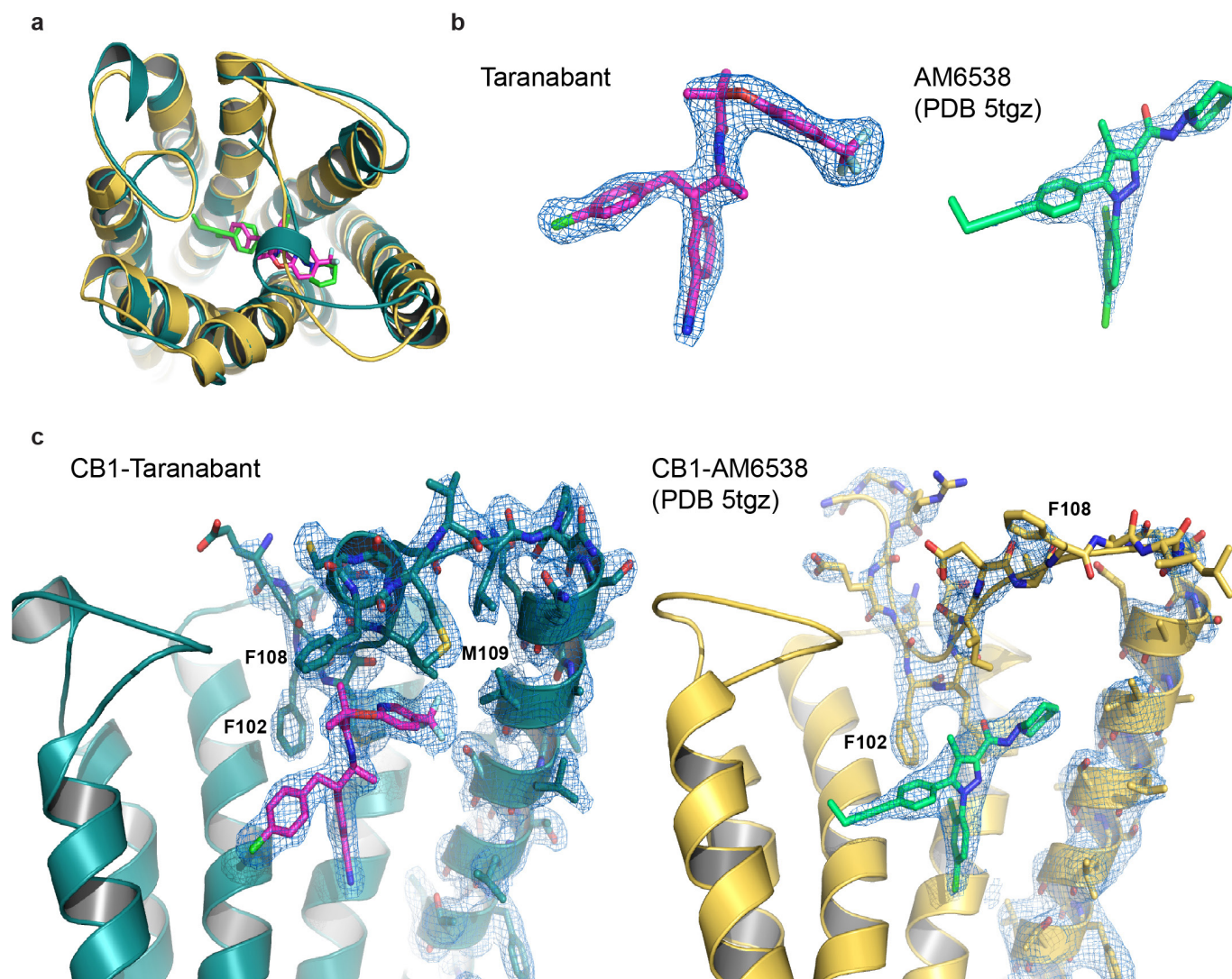


Extended Data Figure 7 | Molecular dynamics simulation of the CB1 structure. **a**, A 60-ns molecular dynamics simulation of the CB1 receptor (after removing the PGS fusion protein) with taranabant present. Black trace is for the entire receptor, red trace is for only the structured

membrane-proximal N-terminal region. **b**, 60-ns molecular dynamics simulation of the CB1 receptor without a ligand present. Black trace is for the entire receptor, red trace is for only the structured membrane-proximal N-terminal region.

Human_CB1	-----MKSILDGLADTTFRITITDLLYVGSNDIQYEDIKGMASKLGYFPQKFPLT	51
Mouse_CB1	-----MKSILDGLADTTFRITITDLLYVGSNDIQYEDIKGMASKLGYFPQKFPLT	51
Chicken_CB1	-----MKSILDGLADTTFRITITDLLYVGSNDIQYEDMKGMASKLGYFPQKFPLS	51
Zebrafish_CB1	MLFPASKSDVKSVDLGVAAETTFRTITSGLQYIGSNDIGYDDHIIDGDFSKSGYPLPKPFA	60
Human_CB2	-----	0
Human_CB1	SFRGSPFQEKMTAGDNPQLV-PA-DQVNITEFYNKSLSSFKENEENIQCGENFMDIECFM	109
Mouse_CB1	SFRGSPFQEKMTAGDNSPLV-PAGDTTNITEFYNKSLSSFKENEDNIQCGENFMDMECFM	110
Chicken_CB1	SFRGDPFQEKMTAGDDPLLSIIPSDQINITEFYNKSLSTFKENEENIQCGENFMDMECFM	111
Zebrafish_CB1	AYRRSSFADKVAPDEELIVKGLPFYPTNSSDVFGN---WSHAEDGSLQCGENFMDMECFM	117
Human_CB2	-----MEECWVTEIANGSKDGLDNPMPKLYM	26
	. : * : : : *	
	TM1 TM2	
Human_CB1	VLNPSQQLATAVISLTLTGTFVLENLLVLCVILHSRSLRCRPSYHFIGSLAVADLLGSVI	169
Mouse_CB1	ILNPSQQLATAVISLTLTGTFVLENLLVLCVILHSRSLRCRPSYHFIGSLAVADLLGSVI	170
Chicken_CB1	ILNPSQQLATAVISLTLTGTFVLENLLVLCVILHSRSLRCRPSYHFIGSLAVADLLGSVI	171
Zebrafish_CB1	ILTPSQQLATAVISLTLTGTFVLENLVLCVILQSRTRLRCRPSYHFIGSLAIADLLGSVI	177
Human_CB2	ILSGPQKTAVAVLCITLLGLLSALENAVLYLILSSHQLRRKPSYLFISLAGADFLASVV	86
	*. * : * : * : * : * : * : * : * : * : * : * : * : * : * : *	
	TM2 TM3	
Human_CB1	FVYSFIDFHVHFRKDSRNVFLFKLGGVTASFTASVGSFLTAIDRYISIHRLPLAYKRIVT	229
Mouse_CB1	FVYSFVDFHVHFRKDSPNVFLFKLGGVTASFTASVGSFLTAIDRYISIHRLPLAYKRIVT	230
Chicken_CB1	FVYSFVDFHVHFRKDSPNVFLFKLGGVTASFTASVGSFLTAIDRYISIHRLPLAYKRIVT	231
Zebrafish_CB1	FVYSFLDFHVHFRKDSPNVFLFKLGGVTASFTASVGSFLTAIDRYVSIHRLPSYRRIVT	237
Human_CB2	FACSFVNFHVHFGVDSKAVFLKIGSVTMTFTASVGSLLTAIDRYLCLRYPPSYKALLT	146
	*. * : * : * : * : * : * : * : * : * : * : * : * : * : * : *	
	TM4 TM5	
Human_CB1	RPKAVVAFCLMWTIAIVIAVLPLLGNWCKLQSVCSDFPHIDETYLMEFWIGVTSVLLLF	289
Mouse_CB1	RPKAVVAFCLMWTIAIVIAVLPLLGNWCKLQSVCSDFPLIDETYLMEFWIGVTSVLLLF	290
Chicken_CB1	RPKAVVAFVCMWTIAIVIAVLPLLGNWCKLNSVCSDFPLIDETYLMEFWIGVTSVLLLF	291
Zebrafish_CB1	RTKAVIAFCMMWAISIIIAVLPLLGNWCKRLNSVCSDFPLIDENYLMFWIGVTSVLVLF	297
Human_CB2	RGRALVTLGIMWVLSALVSYPLMGWTCPP--RPCSELEPLIPNDYLLSWLLFIAFLFSG	204
	* : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *	
	TM5 TM6	
Human_CB1	IVYAYMYILWKAHSHAVRMIQRGTQKSIIHTSEDGKVQVTRPDQARMDIRLAKTLVLIL	349
Mouse_CB1	IVYAYMYILWKAHSHAVRMIQRGTQKSIIHTSEDGKVQVTRPDQARMDIRLAKTLVLIL	350
Chicken_CB1	IVYAYMYILWKAHSHAVRMLQRGTQKSIIQSTEDGKVQITRPDQTRMDIRLAKTLVLIL	351
Zebrafish_CB1	IIYAYMYILWKAHHAVRMLRRTSQKSLVHVSADGTVQVTRPDQARMDIRLAKTLVLIL	357
Human_CB2	IITYYGHVLWKAHQHVASLSG-----HQDRQVPGMARMRLDVRLAKTLGLVL	251
	* : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *	
	TM6 TM7	
Human_CB1	VVLIICWGPPLAIVYDVFQGMNKLIKTVFAFCSMCLLNSTVNPPIIYALRSKDLRHAFR	409
Mouse_CB1	VVLIICWGPPLAIVYDVFQGMNKLIKTVFAFCSMCLLNSTVNPPIIYALRSKDLRHAFR	410
Chicken_CB1	VVLIICWGPPLAIVYDVFQGMNKLIKTVFAFCSMCLLNSTVNPPIIYALRSKDLRHAFR	411
Zebrafish_CB1	VVLVICWGPPLAIVYDLFWRMGDNIKTVFAFCSMCLLNSTVNPPIIYALRSKDLRR AFL	417
Human_CB2	AVLLICWFPVLALMAHSLATLSDQVKKAFACFCSMCLINSMVNPVIYALRSGEIRSSAH	311
	. * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *	
Human_CB1	SMF---PSCEGTAQ-----P---LDNSMGDSCLHKKHANNAASV	442
Mouse_CB1	SMF---PSCEGTAQ-----P---LDNSMGDSCLHKKHANNTASM	443
Chicken_CB1	SMF---PTCEGTAQ-----P---LDNSM-ESDCQKHANNAGNV	443
Zebrafish_CB1	AAC---QGCRGTST-----TPLQLDNSL-ESDC-----HRNQ	445
Human_CB2	HCLAHWKCVRLGSEAKEEAPRSSVTETEDGKITPWPDSRDLDLSDC-----	360
	* . . . : * : *	
Human_CB1	HRAAESCIKSTVKIAKVTMSVSTDTSAEAL	472
Mouse_CB1	HRAAESCIKSTVKIAKVTMSVSTDTSAEAL	473
Chicken_CB1	HRAAESCIKSTVKIAKVTMSVSTDTTAEAL	473
Zebrafish_CB1	HRAAESCVKTTVKIAKLTVMSVSAETSAEAV	475
Human_CB2	-----	360

Extended Data Figure 8 | Sequence alignment of the entire sequence of CB1 from several different species, along with human CB2. The blue boxes denote positions that make contact with taranabli within a 4 Å cut-off. The alignment was performed using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).



Extended Data Figure 9 | Comparison of the structures of CB1 bound to taranabant and CB1 bound to AM6538 (ref. 34; PDB accession 5TGZ).

a, Superposition of the two CB1 structures viewed from the extracellular space. The taranabant-bound structure is shown as a teal cartoon (ligand as magenta sticks), while the AM6538-bound structure is shown as a gold cartoon (ligand as green sticks). **b**, Comparison of $2F_o - F_c$ electron density (contoured at 1.5σ) for the ligands in each structure. On the left is

taranabant from the current structure, on the right is AM6538 from ref. 34. **c**, Comparison of the membrane-proximal N-terminal regions in each structure. On the left is a side view of CB1 from the current structure, with $2F_o - F_c$ electron density (contoured at 1.0σ) shown for the N-terminal region, TM1, and taranabant. On the right is the analogous side view of CB1 from ref. 34 (gold cartoon), with $2F_o - F_c$ electron density (contoured at 1.0σ) shown for the N-terminal region, TM1 and AM6538.

Extended Data Table 1 | Data collection and refinement statistics

	CB1-PGS with Taranabant [†]
Data collection	
Space group	P2 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	50.7, 80.4, 81.2
β (°)	91.7
Resolution (Å)	50.00-2.60 (2.69-2.60) [‡]
<i>R</i> _{sym} or <i>R</i> _{merge} [*]	0.19 (NA)
<i>I</i> / σ <i>I</i>	7.43 (0.96)
Completeness (%)	96.8 (96.9)
Redundancy	5.4 (5.1)
<i>CC</i> _{1/2} in highest shell	0.69
Refinement	
Resolution (Å)	50-2.60
No. reflections	11084
<i>R</i> _{work} / <i>R</i> _{free}	0.19 / 0.23
No. atoms	
Protein	3762
Ligand/ion	56
Other (Lipid and water)	125
B-factors	
Receptor	45.5
Fusion protein	38.3
Ligand	42.0
Ion	91.0
Other (Lipid and water)	44.4
R.m.s deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.20

^{*}Diffraction data from 42 crystals were merged into a single dataset.[†]Values in parentheses are for the highest-resolution shell.[‡]*R*_{merge} > 1 is statistically meaningless, Scalepack⁴⁰ does not report it.

Structure of RNA polymerase I transcribing ribosomal DNA genes

Simon Neyer^{1*}, Michael Kunz^{2*}, Christian Geiss², Merle Hantsche¹, Victor-Valentin Hodirnau², Anja Seybert², Christoph Engel¹, Margot P. Scheffer², Patrick Cramer^{1§} & Achilleas S. Frangakis^{2§}

RNA polymerase I (Pol I) is a highly processive enzyme that transcribes ribosomal DNA (rDNA) and regulates growth of eukaryotic cells^{1–4}. Crystal structures of free Pol I from the yeast *Saccharomyces cerevisiae* have revealed dimers of the enzyme stabilized by a ‘connector’ element and an expanded cleft containing the active centre in an inactive conformation^{5–7}. The central bridge helix was unfolded and a Pol-I-specific ‘expander’ element occupied the DNA-template-binding site. The structure of Pol I in its active transcribing conformation has yet to be determined, whereas structures of Pol II and Pol III have been solved with bound DNA template and RNA transcript^{8–10}. Here we report structures of active transcribing Pol I from yeast solved by two different cryo-electron microscopy approaches. A single-particle structure at 3.8 Å resolution reveals a contracted active centre cleft with bound DNA and RNA, and a narrowed pore beneath the active site that no longer holds the RNA-cleavage-stimulating domain of subunit A12.2. A structure at 29 Å resolution that was determined from cryo-electron tomograms of Pol I enzymes transcribing cellular rDNA confirms contraction of the cleft and reveals that incoming and exiting rDNA enclose an angle of around 150°. The structures suggest a model for the regulation of transcription elongation in which contracted and expanded polymerase conformations are associated with active and inactive states, respectively.

The structure of Pol I, which consists of 14 subunits with a total molecular weight of 590 kDa, has been previously described in atomic detail in an inactive conformation using X-ray crystallography^{5–7}. To determine the structure of transcribing Pol I, we performed single-particle cryo-electron microscopy (single-particle cryo-EM) with a reconstituted yeast Pol I elongation complex containing a DNA–RNA scaffold (Fig. 1a, Extended Data Fig. 1), similar to that used to study transcribing mammalian Pol II¹¹. Particle classification enabled us to reconstruct the Pol I elongation complex structure at 3.8 Å resolution from approximately 94,000 single particles (Fig. 1c, Extended Data Figs 2, 3). The electron density revealed the downstream DNA, the DNA–RNA hybrid (Fig. 1b), and all Pol I domains with the exception of the flexibly linked C-terminal domain of subunit A49 (refs 12, 13) and the C-terminal domain of subunit A12.2. An atomic model was obtained by fitting rigid domains of the Pol I crystal structure⁵, positioning nucleic acids from the bovine Pol II elongation complex structure¹¹, and manually rebuilding regions that were structurally altered (Extended Data Table 1).

Comparison of the resulting Pol I elongation complex structure with the previous Pol I structure^{5,6} reveals that the active centre cleft is contracted by up to 13 Å (Fig. 2a). Contraction occurs through relative movement of the two major polymerase modules ‘core’ and ‘shelf’¹⁴, as predicted⁵. The shelf module moves together with the clamp domain as a single ‘shelf-clamp’ unit, slightly rotating with respect to the core module (Fig. 2e, Supplementary Video 1). Another module, the

‘jaw-lobe’, moves closer to downstream DNA by up to 7 Å (Fig. 2a, d). Comparison of the Pol I elongation complex with elongation complex structures of Pol II^{8,9}, Pol III¹⁰, and bacterial RNA polymerase^{15,16} reveal that all of these polymerases adopt a similar contracted conformation in their transcribing state and underscores the fundamental structural and mechanistic similarity of cellular RNA polymerases from bacteria to eukaryotes (Extended Data Fig. 4).

In the elongation complex structure, the connector is detached from Pol I, as observed when Pol I is bound to the initiation factor Rrn3 (ref. 13, 17). The expander is also displaced, enabling Pol I to form extensive interactions with the DNA–RNA hybrid (Fig. 1a). The enzyme contacts the DNA template at positions +4 to –9 and the RNA transcript at positions –1 to –8 (+1 represents the nucleotide addition site). Pol I generally binds nucleic acids with the same elements as Pol II⁹, but uses several Pol-I-specific residues to contact the upstream part of the DNA–RNA hybrid. The active centre adopts a catalytically competent conformation. The bridge helix is folded throughout (Fig. 2b, c). The trigger loop has weaker electron density, indicating higher mobility. The tip of the trigger loop lacks density, allowing for binding of the nucleoside triphosphate substrate. The polymerase switch regions and cleft loops adopt similar positions as in the Pol II elongation complex⁹ except that fork loop 1 is bent away from the hybrid (Extended Data Fig. 5a), as in the Pol III elongation complex¹⁰ and in a Pol II initiation intermediate¹⁸.

The Pol I elongation complex structure also provides insights into the regulation of the intrinsic RNA cleavage activity of Pol I. RNA cleavage requires subunit A12.2 (refs 19, 20), which consists of two domains. The N-terminal domain resembles that of the Pol II subunit Rpb9, whereas the C-terminal domain corresponds to the catalytic domain of the Pol II RNA cleavage factor TFIIS^{21,22}. In the elongation complex structure, the N-terminal domain of A12.2 remains at the outer rim of the Pol I funnel region, whereas its C-terminal domain is displaced from the pore that it occupies in the Pol I crystal structures^{5,6,17}. Displacement of the A12.2 C-terminal domain from the pore apparently occurs during cleft contraction because modelling of this domain in the pore results in a clash with the contracted shelf module (Extended Data Fig. 5b, c). Thus A12.2 can only enter the active centre when the cleft is fully or partially expanded. This predicts that Pol I adopts a partially expanded conformation during A12.2 action, which is required for RNA proof-reading and polymerase reactivation after backtracking.

To investigate whether the structural differences between the Pol I elongation complex and the free Pol I dimer arise from nucleic acid binding or from conversion of a dimer to a monomer, we also solved the structure of monomeric Pol I in the absence of nucleic acids at 4.0 Å resolution using approximately 80,000 single particles from the same data set (Extended Data Figs 2, 6a, Methods). In this structure, the connector and expander were also displaced, but the cleft was only partially contracted, similarly to the Pol-I–Rrn3 complex^{13,17} (Extended

¹Max-Planck-Institute for Biophysical Chemistry, Department of Molecular Biology, Am Fassberg 11, 37077 Göttingen, Germany. ²Buchmann Institute for Molecular Life Sciences and Institute for Biophysics, Goethe University Frankfurt, Max-von-Laue Str. 15, Frankfurt 60438, Germany.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

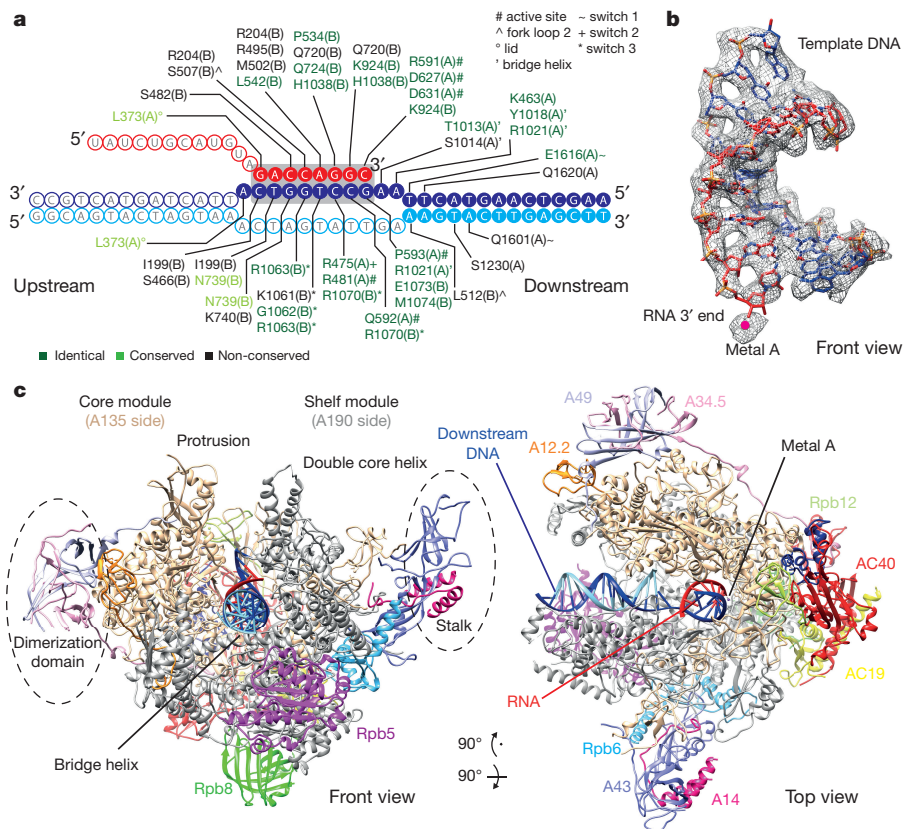


Figure 1 | Single-particle cryo-EM structure of yeast Pol I elongation complex at 3.8 Å resolution. **a**, Nucleic acid scaffold and interactions between Pol I and nucleic acid. Template DNA, non-template DNA and RNA are shown in medium blue, sky blue, and red, respectively. Filled circles represent nucleotides that were well resolved in the electron density. Pol I residues within 4 Å distance are depicted together with the subunit identifier (A for A190, B for A135). **b**, Electron density for the DNA-RNA hybrid with the final model superimposed. The active site metal ion A

is depicted as a magenta sphere. **c**, Ribbon model of the Pol I elongation complex. The view is from the 'front'¹⁴ with the incoming downstream DNA pointing towards the reader. The colouring of the surfaces is according to the standard polymerase subunit colouring: A190, grey; A135, wheat; A49, light blue; A43, slate; AC40, red; A34.5, pink; Rpb5, magenta; Rpb6, silver-blue; AC19, yellow; Rpb8, green; A14, hot pink; A12.2, orange; Rpb10, blue; Rpb12, lemon. Template DNA, non-template DNA and RNA are depicted in medium blue, sky blue and red, respectively.

Data Fig. 6b, c). The central bridge helix remained partially unwound, and the C-terminal domain of A12.2 remained in the pore (Extended Data Fig. 6d–f), confirming that the partially expanded conformation is required for A12.2-dependent RNA cleavage.

Thus, conversion of the Pol I dimer to a monomer leads to a partially expanded conformation, but not to the fully contracted active conformation. The partially expanded conformation resembles the conformation observed when the enzyme adopts a paused²³ or an inhibited²⁴ state (Extended Data Fig. 6g). In both, the bacterial polymerase and Pol I, movement of a rigid shelf-clamp unit away from the core module allows for expansion of the cleft and a coordinated widening of the pore (called the 'secondary channel' in bacterial RNA polymerase). This movement involves a slight rotation of the shelf-clamp unit with respect to the core module, reflected in the term 'ratcheting' used in one study of the bacterial polymerase²⁴.

Available data thus suggest that RNA polymerases can adopt partially expanded and contracted conformations that are associated with inactive and active states, respectively. Binding of nucleic acids in the cleft apparently maintains the contracted conformation and excludes A12.2 from the pore, whereas rearrangements in the nucleic acids upon misincorporation or pausing could induce the partially expanded conformation that is transcriptionally inactive but enables A12.2 entry into the pore and enzyme reactivation by RNA cleavage. According to this model, transcription elongation can be regulated by allosteric coupling of nucleic acid binding with cleavage factor binding in the cleft and pore, respectively, via contraction and expansion of the polymerase.

To investigate the physiological relevance of the single-particle cryo-EM structure, we further determined the structure of the natural

Pol I elongation complex that forms in yeast cells by promoter-dependent initiation on rDNA with the use of cryo-electron tomography (Fig. 3). We spread active rDNA genes from exponentially growing yeast cells onto an electron microscopy grid such that they formed 'Miller trees'²⁵ (Extended Data Fig. 7). To overcome previous limitations in sample preparation, we used instant plunge-freezing to keep the sample in a close-to-native environment. The obtained images revealed the detailed arrangement of Pol I enzymes along rDNA, nascent RNA emerging from Pol I, and large densities at the RNA ends that resemble 5' classical knobs²⁶ (Fig. 3a, b). From the cryo-electron tomography images, we selected 11 complete Miller trees and several smaller Pol I trails, each containing 10–20 Pol I enzymes with associated RNA. This yielded 993 transcribing Pol I enzymes for further analysis.

We observed that each rDNA gene is loaded with ~70 Pol I enzymes, which showed a median centre-to-centre distance of 18 ± 10 nm (Fig. 3c, Extended Data Fig. 8a), consistent with previous results²⁶. Only ~2% of the Pol I complexes were separated by a distance of less than 12 nm, which would allow for interaction between enzymes. Furthermore, consecutive enzymes show random relative orientations, arguing against specific interactions that were suggested previously²⁷. The incoming and exiting rDNA enclose an angle of ~150° measured at each triple of successive Pol I molecules (Fig. 3c, Extended Data Fig. 8b). This angle was independent of the length of the DNA between enzymes and could not be obtained from the single-particle cryo-EM structure, because density for upstream DNA was poor.

After classification we performed sub-tomogram averaging ($n = 225$) to obtain a cryo-electron tomography structure of the cellular transcribing Pol I at a resolution of ~29 Å (Fourier shell correlation (FSC) 0.5

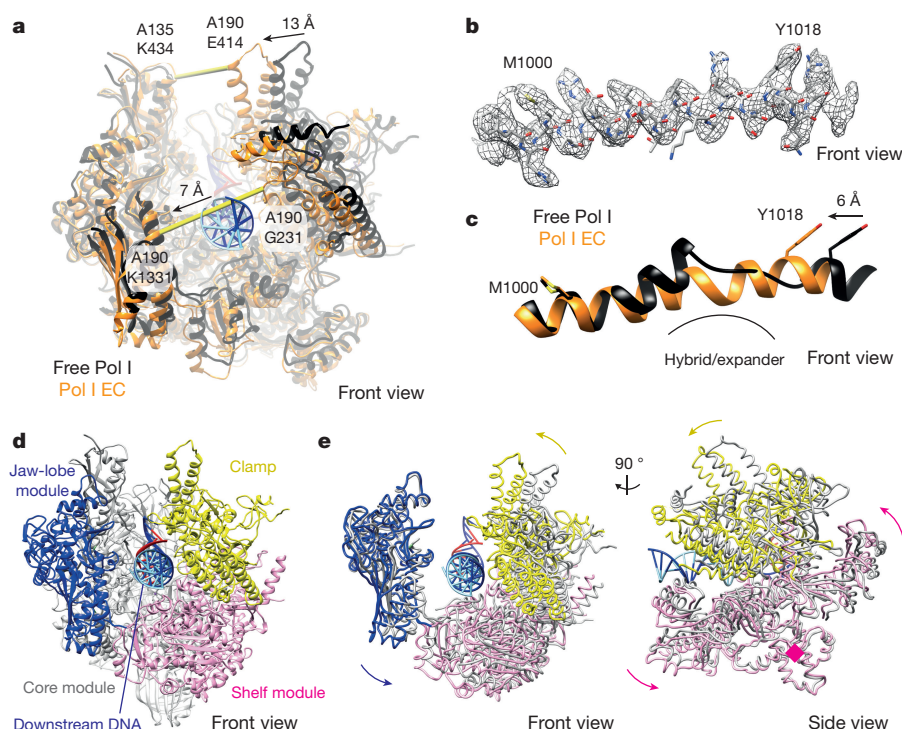


Figure 2 | Cleft contraction and module movements. **a**, Comparison of structures of Pol I elongation complex (EC) (orange) and free dimeric Pol I (PDB: 4C2M; black) after superposition of their A135 subunits. Cleft width was measured between subunit A190 residue E414 and subunit A135 residue K434. For clarity, only subunits A190 and A135 are displayed. **b**, Electron density of the folded bridge helix in the Pol I elongation complex. **c**, Comparison of bridge helices in the elongation complex (orange) and free Pol I (black). **d**, Pol I elongation complex ribbon model coloured by four mobile modules. The peripheral subcomplexes A14–A43 and A49–A34.5 are omitted for clarity. **e**, Movements of polymerase modules upon cleft contraction. Ribbon models of free Pol I (grey) and elongation complex are shown after superposition of their core modules (omitted). Arrows indicate movement and rotation of the clamp-shelf and the jaw-lobe modules. Colour code as in Fig. 1.

criterion; ~ 25 Å with the FSC 0.143 criterion; Extended Data Fig. 9a). The cryo-electron tomography structure strongly resembled the single-particle cryo-EM structure, showing an overall cross-correlation

score of 0.85. An FSC plot between the cryo-electron tomography and single-particle cryo-EM structures decreased beyond the 0.143 threshold at 31 Å (Extended Data Fig. 9a). The peripheral

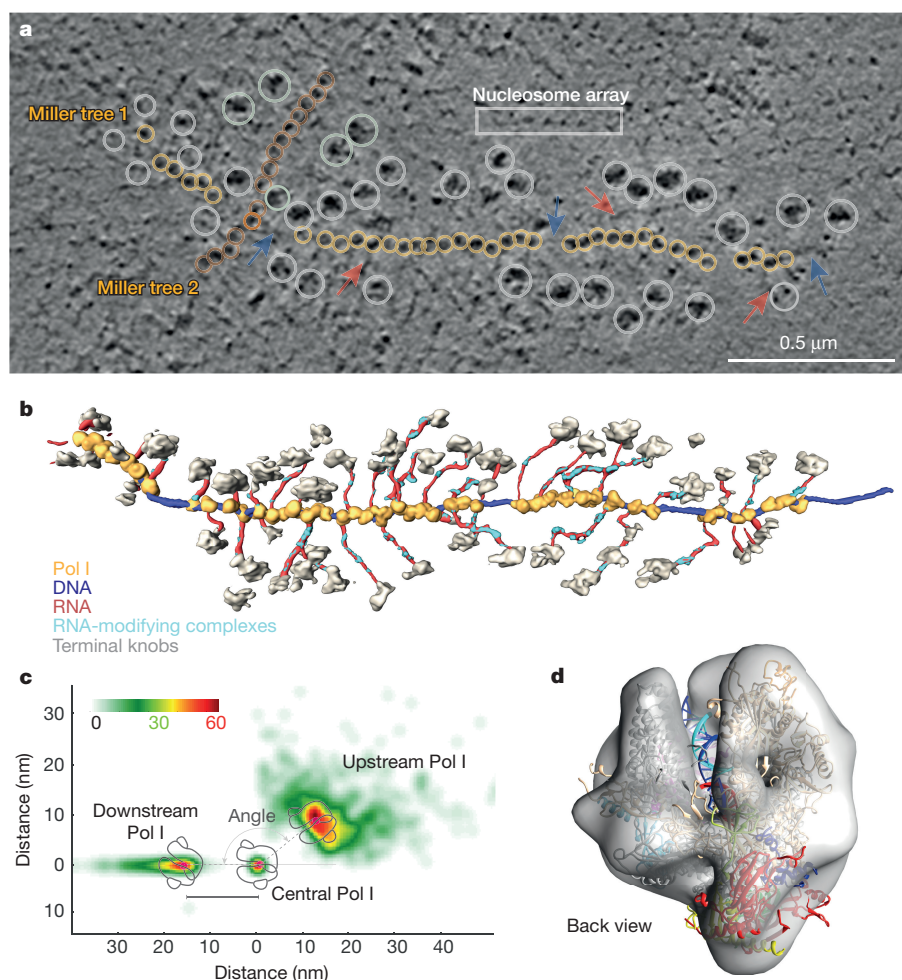


Figure 3 | Cryo-electron tomography analysis of Pol I transcribing rDNA genes. **a**, 2-nm thick tomographic slice through a cryo-electron tomography image with two of the Miller trees, showing the terminal knobs (grey circles), the DNA (typical examples marked by blue arrows), the RNA (red-pink arrows), and the Pol I enzymes (yellow and dark yellow circles for Miller trees 1 and 2, respectively). Several nucleosomes are attached to DNA like beads on a string (white box). **b**, Three-dimensional surface rendering of Miller tree 1 in **a**, showing the terminal knobs (light grey), DNA (blue), RNA (red), possible RNA-modifying complexes (cyan), and Pol I complexes (yellow). **c**, Schematic of three successive Pol I enzymes (of the upstream Pol I, the central Pol I, and the downstream Pol I) overlaid with their probability density localization (heat map) (Extended Data Fig. 8a, b). **d**, Fit of the Pol I elongation complex ribbon model from single-particle cryo-EM into the cryo-electron tomography reconstruction in grey. The good fit observed here is not possible with the expanded conformation of Pol I (Extended Data Fig. 9b). Colour code as in Fig. 1.

subcomplexes A14–A43 and A34.5–A49 were flexible, consistent with the weaker density observed in the single-particle cryo-EM structure (Extended Data Fig. 2). The width of the active centre cleft was the same in both structures (Fig. 3d), confirming that the contracted single-particle cryo-EM structure represented the natural conformation of actively transcribing Pol I (Extended Data Fig. 9b).

Taken together, we used here two independent cryo-electron microscopic approaches to define the contracted Pol I conformation as the active transcribing state of the enzyme. This revealed that all three eukaryotic RNA polymerases adopt a highly similar closed active centre conformation during transcription elongation. Together with published data, our results provide evidence that the elongation phase of transcription is regulated by cleft contraction and expansion. In particular, rearrangements of nucleic acids in the cleft above the active site cleft may be coupled to binding of factors in the pore beneath the cleft. Thus Pol I does not only undergo induced fit to align nucleic acids with the catalytic site, it is apparently also regulated allosterically.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 August; accepted 25 October 2016.

Published online 14 November 2016.

1. Roeder, R. G. & Rutter, W. J. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* **224**, 234–237 (1969).
2. Novello, F. & Stirpe, F. Simultaneous assay of RNA polymerase I and II in nuclei isolated from resting and growing rat liver with the use of alpha-amanitin. *FEBS Lett.* **8**, 57–60 (1970).
3. Grummt, I. The nucleolus—guardian of cellular homeostasis and genome integrity. *Chromosoma* **122**, 487–497 (2013).
4. Goodfellow, S. J. & Zomerijk, J. C. Basic mechanisms in RNA polymerase I transcription of the ribosomal RNA genes. *Subcell. Biochem.* **61**, 211–236 (2013).
5. Engel, C., Sainsbury, S., Cheung, A. C., Kostrewa, D. & Cramer, P. RNA polymerase I structure and transcription regulation. *Nature* **502**, 650–655 (2013).
6. Fernández-Tornero, C. *et al.* Crystal structure of the 14-subunit RNA polymerase I. *Nature* **502**, 644–649 (2013).
7. Kostrewa, D., Kuhn, C. D., Engel, C. & Cramer, P. An alternative RNA polymerase I structure reveals a dimer hinge. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 1850–1855 (2015).
8. Gnat, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 (2001).
9. Kettenberger, H., Armache, K. J. & Cramer, P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell* **16**, 955–965 (2004).
10. Hoffmann, N. A. *et al.* Molecular structures of unbound and transcribing RNA polymerase III. *Nature* **528**, 231–236 (2015).
11. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M. & Cramer, P. Structure of transcribing mammalian RNA polymerase II. *Nature* **529**, 551–554 (2016).
12. Jennebach, S., Herzog, F., Aebersold, R. & Cramer, P. Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage. *Nucleic Acids Res.* **40**, 5591–5601 (2012).
13. Pils, M. *et al.* Structure of the initiation-competent RNA polymerase I and its implication for transcription. *Nat. Commun.* **7**, 12126 (2016).
14. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* **292**, 1863–1876 (2001).

15. Vassilyev, D. G., Vassilyeva, M. N., Perederina, A., Tahirov, T. H. & Artsimovitch, I. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448**, 157–162 (2007).
16. Sekine, S., Murayama, Y., Svetlov, V., Nudler, E. & Yokoyama, S. The ratcheted and ratchetable structural states of RNA polymerase underlie multiple transcriptional functions. *Mol. Cell* **57**, 408–421 (2015).
17. Engel, C., Plitzko, J. & Cramer, P. RNA polymerase I-Rrn3 complex at 4.8 Å resolution. *Nat. Commun.* **7**, 12129 (2016).
18. Plaschka, C. *et al.* Transcription initiation complex structures elucidate DNA opening. *Nature* **533**, 353–358 (2016).
19. Sosunov, V. *et al.* Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *EMBO J.* **22**, 2234–2244 (2003).
20. Kuhn, C. D. *et al.* Functional architecture of RNA polymerase I. *Cell* **131**, 1260–1272 (2007).
21. Kettenberger, H., Armache, K. J. & Cramer, P. Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* **114**, 347–357 (2003).
22. Ruan, W., Lehmann, E., Thomm, M., Kostrewa, D. & Cramer, P. Evolution of two modes of intrinsic RNA polymerase transcript cleavage. *J. Biol. Chem.* **286**, 18701–18707 (2011).
23. Weixlbaumer, A., Leon, K., Landick, R. & Darst, S. A. Structural basis of transcriptional pausing in bacteria. *Cell* **152**, 431–441 (2013).
24. Tagami, S. *et al.* Crystal structure of bacterial RNA polymerase bound with a transcription inhibitor protein. *Nature* **468**, 978–982 (2010).
25. Miller, O. L., Jr & Beatty, B. R. Visualization of nucleolar genes. *Science* **164**, 955–957 (1969).
26. Osheim, Y. N. *et al.* Pre-18S ribosomal RNA is structurally compacted into the SSU processome prior to being cleaved from nascent transcripts in *Saccharomyces cerevisiae*. *Mol. Cell* **16**, 943–954 (2004).
27. Albert, B. *et al.* RNA polymerase I-specific subunits promote polymerase clustering to enhance the rRNA gene transcription cycle. *J. Cell Biol.* **192**, 277–293 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank O. Gadal and I. Léger-Silvestre for technical assistance with the yeast Miller tree spreading technique, and H. Schwalbe for initial discussions. We thank T. Gubbey for initial experiments on the Pol I elongation complex and C. Bernecky, C. Plaschka and D. Tegunov for support with the single-particle data analysis. We thank T. Schulz for yeast fermentation. S.N. was supported by a PhD student fellowship from the Boehringer Ingelheim Fonds. P.C. was supported by the Deutsche Forschungsgemeinschaft (SFB860, SPP1935), the Advanced Grant ‘TRANSREGULON’ from the European Research Council (grant agreement No 693023), and the Volkswagen Foundation. A.S.F. was supported by the Deutsche Forschungsgemeinschaft (SFB 902) and the Starting Grant ‘JTOMO’ from the European Research Council.

Author Contributions S.N. planned and carried out the single particle sample preparation, data collection and data analysis. M.K. planned and carried out the tomographic data analysis. C.G. carried out the sample preparation for tomography. M.H. advised on structure determination procedures. V.V.H. advised on and carried out sample preparation for tomography. A.S. advised on sample preparation for tomography. C.E. advised on biochemical procedures. M.P.S. advised on tomographic data analysis. P.C. designed and supervised research, and supervised single particle structure determination. A.S.F. designed and supervised research, supervised single particle data collection and performed tomographic data collection and analysis. S.N., P.C. and A.S.F. prepared the manuscript, with contributions from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.S.F. (achilleas.frangakis@biophysik.org) and P.C. (patrick.cramer@mpibpc.mpg.de).

Reviewer Information *Nature* thanks R. Ebright, E. Nogales and E. Nudler for their contribution to the peer review of this work.

METHODS

Preparation of Pol I elongation complex (EC). Endogenous 14-subunit Pol I was prepared from *Saccharomyces cerevisiae* as described previously⁵, with some modifications. Yeast strain CB010 expressing a C-terminal Flag/10× histidine-tagged A190 subunit was fermented and collected during the exponential phase. For Pol I purification, 350 g of cells were used. Proteins were precipitated overnight at 4°C with ammonium sulphate (2 M). Re-solubilized Pol I was enriched by large-scale affinity purification with Ni-NTA beads (Qiagen). Further enrichment with anion and cation exchange chromatography yielded to pure Pol I enzyme. The sample was applied to a Superose 6 10/300 size-exclusion column (GE Healthcare) in 5 mM HEPES (pH 7.8), 150 mM potassium acetate, 1 mM MgCl₂, 10 μM ZnCl₂ and 10 μM β-mercaptoethanol.

DNA and RNA were purchased from IDT and Exiqon (Vedbaek), respectively. The nucleic acid scaffold sequences were as follows. Template DNA, 5'-AAGC TCAAGTACTTAAGCCTGGTCATTAAGTACTGTC-3'; nontemplate DNA, 5'-GGCAGTACTAGTAACTAGTATTGAAAGTACTTGAGCTT-3'; RNA, 5'-UAUCUGCAUGUAGACCAGGC-3' (in underlined nucleosides, a methylene bridge connects the 2'-O and the 4'-C atoms of the ribose ring, thereby forming locked nucleic acids). Nucleic acids were annealed by continuously decreasing temperature from 95°C to room temperature over a period of 60 min. EC assembly was achieved by incubating Pol I (300 μg, 3.5 mg ml⁻¹) with a twofold molar excess of scaffold for 10 min at room temperature (Extended Data Fig. 1).

Single-particle cryo-EM. For single-particle cryo-EM, Pol I EC complexes at a concentration of 200 μg ml⁻¹ were cross-linked with 0.9 mM BS3 (Sigma Aldrich) for 30 min at 30°C after optimization (Extended Data Fig. 1). The reaction was stopped by adding 50 mM ammonium bicarbonate, and the sample was purified by size-exclusion chromatography on a Superose 6 3.2/300 column (GE Healthcare) equilibrated in 5 mM HEPES (pH 7.8), 150 mM potassium acetate, 1 mM MgCl₂, 10 μM ZnCl₂ and 10 μM β-mercaptoethanol. A 4 μl aliquot of 100 μg ml⁻¹ purified sample was applied to a glow-discharged (10 s) R1.2/1.3 UltraAuFoil grid (Quantifoil), and plunge-frozen in liquid ethane (Vitrobot Mark IV (FEI) at 95% humidity, 4°C, 8.5 s blotting time, blot force 14). Dose-fractionated movies (30 frames, 0.25 s each) were collected at a nominal magnification of 130,000× (1.05 Å per pixel) in nanoprobe energy-filtered transmission electron microscopy (EFTEM) mode at 300 kV with a Titan Krios (FEI) electron microscope using a GIF Quantum s.e. post-column energy filter in zero loss peak mode and a K2 Summit detector (Gatan). The camera was operated in dose-fractionation counting mode with a dose rate of ~7.5 electrons per pixel per second (0.25 s single frame exposure) and a total dose of ~56 electrons per Å². Defocus values ranged from -0.6 to -3 μm with marginal (<0.1 μm) astigmatism. Global motion correction was performed as described²⁸, but single-particle cryo-EM images were not partitioned.

Single-particle cryo-EM image processing. Parameters of the contrast transfer function (CTF) on each micrograph were estimated with CTFIND4 (ref. 29). In a first step, ~1,500 particles were picked with the semi-automated swarm method of EMAN2 e2boxer.py³⁰. Relion was used for the whole-image processing workflow³¹ unless stated otherwise. Reference-free 2D classes were generated, seven of which were used for template-based auto-picking after filtering to 20 Å. We extracted 401,000 particles from 2,300 micrographs with a 230 × 230 pixel box and used them for further processing. Pixels with 5 standard deviations from the mean value were replaced with random values from a Gaussian noise distribution. All images were normalized to make the average density of the background equal to zero during pre-processing. False-positive particles showing very bright dots, which were presumably gold contamination, were removed by manual inspection or unsupervised 2D classification. The remaining 282,000 particles were aligned on a reference generated from the PDB entry 4C2M⁵ filtered to 40 Å. To correct for local motion and for radiation damage, we used the movie processing function of Relion including 'particle polishing', in which the resolution-dependent decay caused by radiation damage is taken into account³¹. Local resolution was estimated as described^{32,33}.

During classification of single-particle cryo-EM images (Extended Data Fig. 2), we first separated out particles lacking nucleic acids. To this end, the Pol I cleft of the average resulting after the first round of alignment was masked. The subsequent classification led to four classes: (1) nucleic acid-free Pol I (115,000 particles); (2) Pol I elongation complex (94,000 particles; hereafter referred to as 'EC'); (3) Pol I elongation complex with an alternative DNA conformation (37,000 particles); and (4) other particles (35,000 particles). Among the nucleic acid-free polymerase particles, 80,000 particles displayed a defined position of the C-terminal domain of A12.2. We refer to the SP average of these particles as the 'monomer'.

In a second step, a mask around the dimerization domain was applied to remove particles from which the A49–34.5 subcomplex dissociated. This led to 32,000 and 40,000 particles in case of the Pol I monomer and Pol I EC, respectively. To visualize the mobile stalk, we then applied a mask around A14/43 during refinement allowing only local searches.

Gold-standard Fourier shell correlations (FSCs) were calculated during the 3D refinement in Relion between two independently refined halves of the data. According to the FSC 0.143 criterion, global resolutions of 4.0 Å and 3.8 Å were estimated for Pol I monomer and EC structures, respectively, which were sharpened with temperature factors of -146 Å² and -149 Å², respectively.

Structural modelling. Two separate models were built for the monomer and the Pol I EC. PDB entry 4C2M⁵ was used as the starting model in both cases. Models were constructed lacking the expander, connector and, in case of the EC, the C-terminal domain of A12.2. The models were further truncated by removing the peripheral subcomplexes A49–34.5 and A14–43. The starting models were placed in densities for the monomer and the EC by fitting in UCSF Chimera³⁴, followed by rigid body fitting with a Phenix real space refinement³⁵. Rigid body groups were defined based on module definitions originally proposed for Pol II¹⁴. A starting model for DNA and RNA was derived from bovine Pol II¹¹ and further refined. Structurally altered regions were adjusted to the density in COOT³⁶ followed by real space refinement in Phenix. To generate complete models, structure of subcomplexes A49–34.5 and A14–43 were fit into the classified map in Chimera. No changes were made within the domains during model building, except for A34.5 C-terminal tail. The models were validated using the FSC between the model and the map, EMRinger³⁷ and Molprobity³⁸.

Miller tree preparation and cryo-electron tomography imaging. Miller chromatin spreads²⁵ were prepared with some modifications as described³⁹, using the NOY1071 yeast strain with 25 copies of ribosomal DNA (rDNA) repeats⁴⁰. Yeast cells were grown to mid-log phase (absorbance (A₆₀₀) = 0.4) in YPG medium supplemented with 1 M sorbitol at 30°C. YPG medium contains 1% (w/v) yeast extract, 2% (w/v) bacto-peptone and 2% glucose. 1 ml yeast cell culture in mid-log phase was added for 4.5 min to the preheated 20T zymolyase (Amsbio, Biotechnology) solution (5 mg/200 μl zymolyase in YPG medium at 30°C) for a slight digestion of the yeast cell wall. Subsequently, the yeast cell culture was centrifuged at 13,000 rpm for 15 s and the pellet was resuspended in 1 ml of 0.0025% Triton-X-100 (Sigma-Aldrich) ddH₂O at pH 9.2 adjusted with pH 10 buffer (Thermo Fischer Scientific). The yeast suspension was transferred to a flask containing 5 ml of 11 mM KCl solution. The lysate was pipetted and incubated in a hydrophobic plastic Petri dish (Carl Roth GmbH & Co. Kg) placed on a shaker for 45 min. Sucrose was excluded from the sucrose-formalin solution as used in ref. 39. To fix chromatin, 400 μl of 37% formaldehyde (Sigma-Aldrich) solution was applied for 5 min. The yeast lysate was deposited on electron microscopy grids with a ~30 nm thick carbon support layer evaporated by a carbon coater 208Carbon (Cressington) and glow discharged for ~1 min using a home-made device. Subsequently they were placed within home-built grid chamber insets and centrifuged within an Eppendorf 5810R centrifuge (Eppendorf) for 5 min at ~2,200g at 20°C. Before plunge-freezing the grids were transferred to an 11 mM KCl solution for which ddH₂O at pH 9.2 was used.

The grids were immediately plunge-frozen in liquid ethane by a Vitrobot Mark IV (FEI) with 25 blotting force, 3 s blotting and 10–15 s draining time and the blotting chamber set to 100% humidity at 10°C. Cryo-grids were mounted into autoloader grids with C-clippings (FEI) in an EM FC6 cryo-microtome (Leica) that was cooled with liquid nitrogen under gaseous flow to -150°C. During mounting, grids were visually inspected to determine whether they contained an intact carbon film.

Tilt-series were recorded using DigitalMicrograph (Gatan Inc.) at a nominal magnification of 33,000× (4.0 Å per pixel) in EFTEM mode at 300 keV using a Titan Krios with a GATAN GIF Quantum s.e. post-column energy filter in zero loss peak mode and a K2 Summit detector. The camera was operated in counting mode with a dose rate of ~15 electrons per pixel per second and a total dose of ~100 electrons per Å². The tilt-series ranged from -63° to +63° with an angular increment of 2° and defocus set at -5 μm. Tilted images were fiducial-less aligned⁴¹ and reconstructed by super-sampling SART⁴². The CTF was measured and corrected in slices in 3D⁴³.

Reconstruction and segmentation of Miller trees. 3D reconstructions were visualized with the EMpackage in Amira (FEI & Zuse Institute)⁴⁴ and analysed in TOM package⁴⁵. Segmentation of the Miller trees was performed manually in Amira by drawing contours encompassing individual features on mildly Gaussian low-pass filtered tomograms using the high-contrast option of super-sampling SART⁴².

Sub-tomogram averaging of Pol I enzymes. Sub-tomograms containing transcribing Pol I enzymes on rDNA were manually selected. The enzymes were re-centred using a Gaussian blob of the size of Pol I. The positions of all enzymes were subsequently indexed such that they were placed sequentially on the DNA. As the DNA was visible in the reconstructions, the indexing was unambiguous (Extended Data Fig. 8d). For sub-tomogram averaging (that is, the cryo-electron tomography structure) we selected five Miller trees according to the following criteria: (1) They visually showed a transcriptional directionality (several Miller trees were

not completely visualized in the field of view). (2) All Pol I enzymes aligned according to the Miller tree directionality. (3) The RNA exit site matched previous observations¹¹ (Extended Data Fig. 8c, e). This resulted in a total of 225 Pol I enzymes contributing to the final sub-tomogram average.

Sub-tomogram averaging was then performed on each Miller tree individually. This was to guarantee that the directionalities of the enzymes were not mixed owing to the globular shape of the enzyme, the pseudo-symmetry axis, and the varying ice thickness of the recording area leading to different signal-to-noise ratio among the enzymes. The Euler angles were determined *a priori* for each of the three consecutive Pol I enzymes per Miller tree by calculating the vector from centre-to-centre position. Constrained sub-tomogram averaging was performed on sub-tomograms with $64 \times 64 \times 64$ voxels using a spherical mask (~ 20 nm diameter). To ensure the robustness of the sub-tomogram averaging, two different starting references were used (1) the average of all rotationally pre-aligned Pol I enzymes per strand, and (2) a Gaussian blob of the size of Pol I. Both converged to approximately the same density. During sub-tomogram averaging of each individual Miller tree, polymerases were low-pass filtered and the alignment was run with a translational freedom of 10 voxels around the Gaussian blob refined position, a full rotational freedom for phi and psi, and a constrained rotational freedom of ± 30 degrees for θ with 5 degrees sampling increment, until the average reached convergence. The missing wedge was taken into account during the entire alignment.

The sub-tomogram averages of each Miller tree were individually inspected and the orientation of the Pol I enzymes on each Miller tree was analysed. The 3' to 5' directionality of the enzymes on each Miller tree was analysed. If all enzymes had the same directionality (that is, the signal-to-noise ratio was sufficient to align them properly), their sub-tomogram average was used for further processing. If the enzymes had conflicting directionalities (including complete random directionality), their sub-tomogram average was rejected. Five Miller trees qualified for this criterion. Their enzyme directionality was visualized compared to the Miller-tree directionality, and they all conformed. Finally, 225 enzymes (from the 993 total enzymes in the tomograms) of the five selected Miller trees were subjected to a refined sub-tomogram averaging and the resulting cryo-electron tomography structure reached a resolution of ~ 29 Å with the FSC 0.5 threshold criterion (~ 31 Å when compared to the single-particle cryo-EM structure).

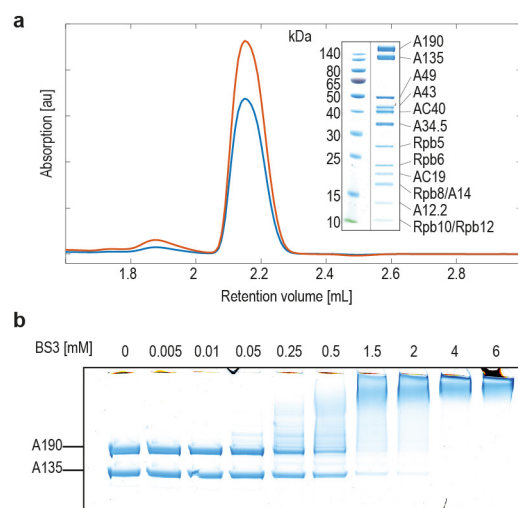
Additional cryo-electron tomography analysis. In the tomograms both the DNA and the RNA could be seen emanating from the enzymes (Extended Data Fig. 8c, d). They were manually localized as close as possible to the enzyme and subsequently sub-tomogram averaging was performed around this position. To obtain evidence for the RNA exit channel visualized in the single-particle cryo-EM structure, we made three independent attempts to manually select the position of exiting RNA on Pol I in the tomogram without prior knowledge of the structure (Extended Data Fig. 8e). The resulting point distribution of exiting RNA on the cryo-electron tomography structure agreed with the location of the RNA exit channel in the single-particle cryo-EM map and further confirmed the correct superposition of the two independent structures.

The distances of consecutive Pol I enzymes were calculated as the Euclidian distance between their centre-to-centre positions. For plotting the probability density function, one enzyme was centred, the downstream enzyme was placed on the x axis, and the upstream enzyme was placed on the plane. Between three consecutive neighbouring enzymes, the in-plane angle was estimated.

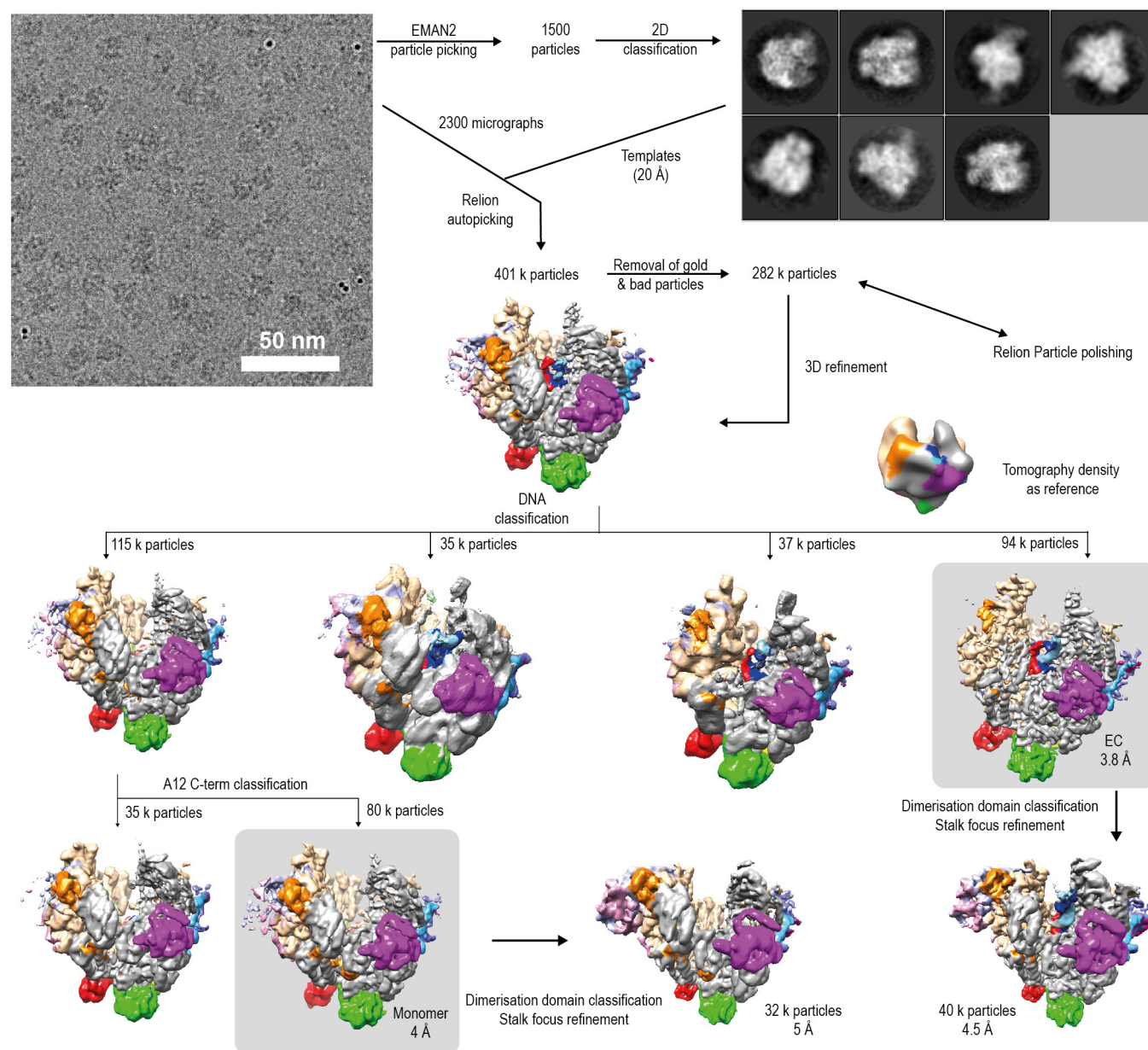
For fitting of structures to the cryo-electron tomography reconstruction, rigid body fitting of the cryo-electron tomography and the single-particle cryo-EM structures of the Pol I EC was performed automatically, using MATLAB scripts (MATLAB and Statistics Toolbox Release 2012b, The MathWorks, Inc., Natick), implemented in the TOM package⁴⁵ (all scripts are freely available upon request) as well as Chimera³⁴. This resulted in a global cross-correlation value of ~ 0.8 and a FSC shown in Extended Data Fig. 9a. The contour level for the cryo-electron tomography structure for volume rendering of our average was calculated from the theoretical molecular mass with an average protein density of 0.8 kDa nm^{-3} .

Data availability statement. Cryo-electron microscopy densities were deposited in the Electron Microscopy Data Base under the accession codes EMD-4147 and EMD-4148 for the EC and the free monomer, respectively. Sub-tomogram average densities were deposited in the Electron Microscopy Data Base under the accession codes EMD-4149. Model coordinates were deposited in the Protein Data Bank under the accession codes 5M3F and 5M3M for the EC and the free monomer, respectively.

28. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
29. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
30. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
31. Scheres, S. H. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012).
32. Plaschka, C. *et al.* Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature* **518**, 376–380 (2015).
33. Cardone, G., Heymann, J. B. & Steven, A. C. One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *J. Struct. Biol.* **184**, 226–236 (2013).
34. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
35. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
36. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
37. Barad, B. A. *et al.* EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).
38. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
39. Osheim, Y. N., French, S. L., Sikes, M. L. & Beyer, A. L. Electron microscope visualization of RNA transcription and processing in *Saccharomyces cerevisiae* by Miller chromatin spreading. *Methods Mol. Biol.* **464**, 55–69 (2009).
40. Machin, F. *et al.* Transcription of ribosomal genes can cause nondisjunction. *J. Cell Biol.* **173**, 893–903 (2006).
41. Castaño-Díez, D., Al-Amoudi, A., Glynn, A. M., Seybert, A. & Frangakis, A. S. Fiducial-less alignment of cryo-sections. *J. Struct. Biol.* **159**, 413–423 (2007).
42. Kunz, M. & Frangakis, A. S. Super-sampling SART with ordered subsets. *J. Struct. Biol.* **188**, 107–115 (2014).
43. Kunz, M. & Frangakis, A. S. Three-dimensional CTF correction improves the resolution of electron tomograms. *J. Struct. Biol.* <http://dx.doi.org/10.1016/j.jsb.2016.06.016> (2016).
44. Pruggnaller, S., Mayr, M. & Frangakis, A. S. A visualization and segmentation toolbox for electron microscopy. *J. Struct. Biol.* **164**, 161–165 (2008).
45. Korinek, A., Beck, F., Baumeister, W., Nickell, S. & Plitzko, J. M. Computer controlled cryo-electron microscopy—TOM² a software package for high-throughput applications. *J. Struct. Biol.* **175**, 394–405 (2011).

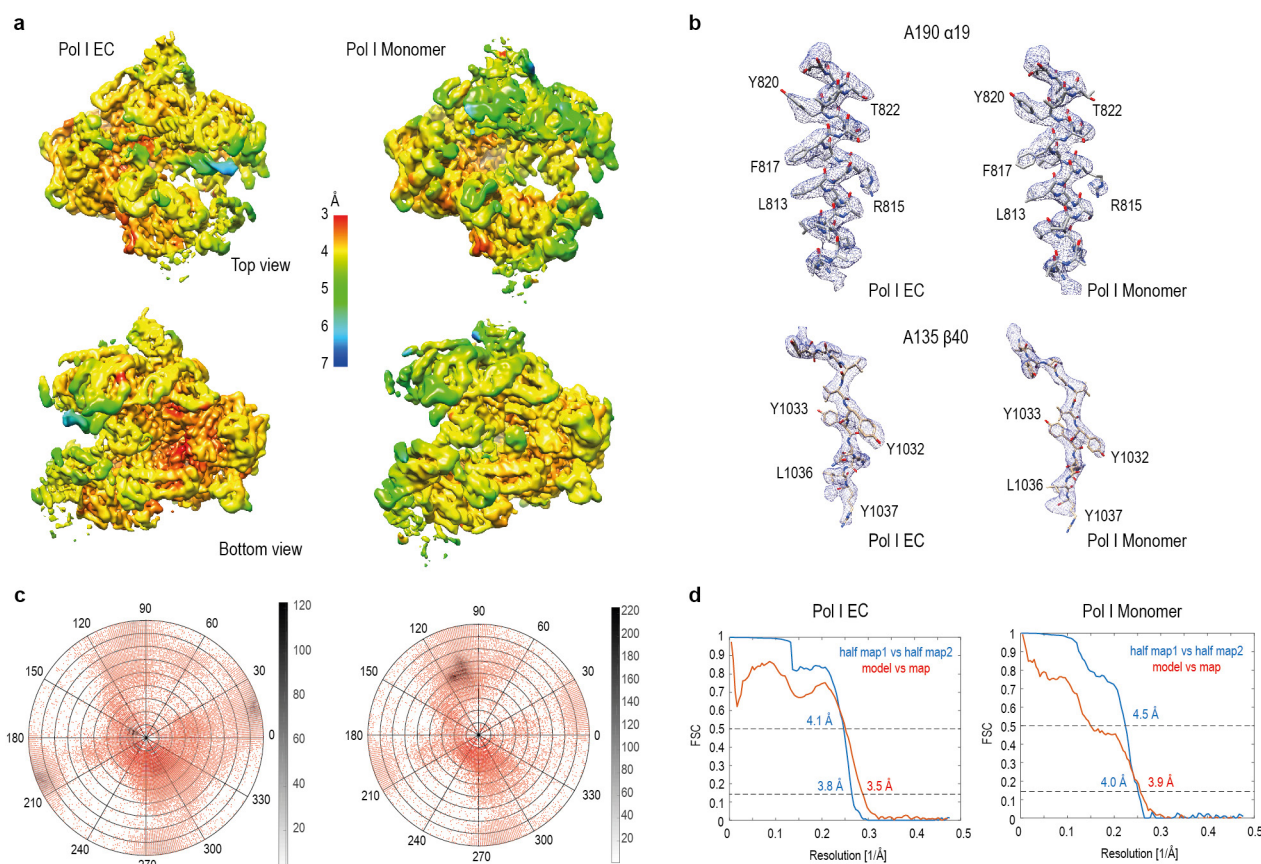


Extended Data Figure 1 | Preparation of Pol I elongation complex (EC) for single-particle cryo-EM. **a**, Size-exclusion chromatogram (Superose 6 Increase 3.2/300; GE Healthcare) of reconstituted Pol I EC. Higher absorbance at 260 nm (red line) than at 280 nm (blue line) indicates presence of nucleic acids. Coomassie-stained SDS-PAGE analysis of pooled peak fractions shows the presence of all 14 Pol I subunits. **b**, Coomassie-stained SDS-PAGE analysis of titration with BS3 cross-linker. Gel is cropped to large subunits A190 and A135. A shift to higher molecular weight is observed with increasing BS3 concentration indicating successful crosslinking. Based on interpolation, we chose 0.9 mM BS3 (not shown) as the appropriate concentration for final sample preparation.



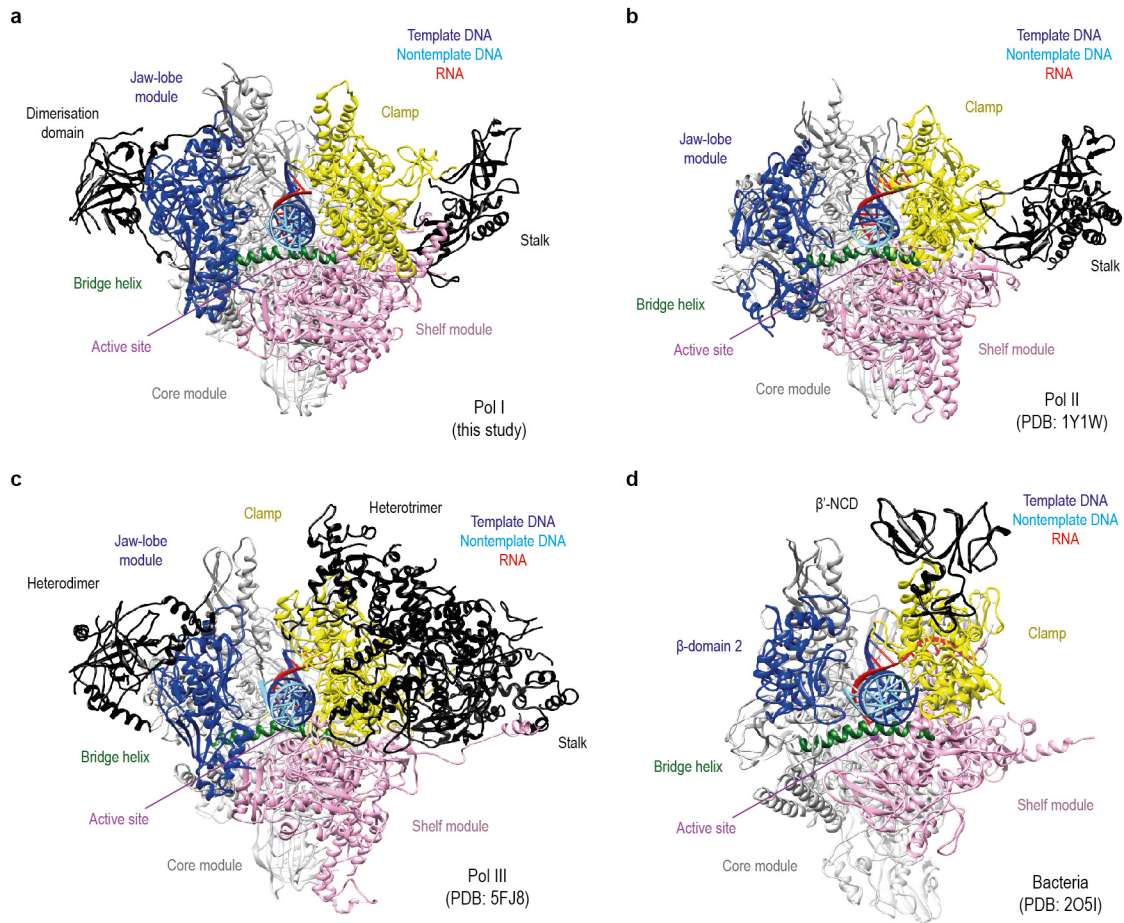
Extended Data Figure 2 | Single-particle cryo-EM particle sorting pipeline. Annotated arrows indicate the direction of processing and provide information regarding the number of particles used and the classification masks applied. A representative micrograph of the Pol I EC under cryo conditions showed particles of the expected size. A set of 1,500 particles was picked manually with EMAN2 (ref. 30) and used to generate initial 2D classes for template based auto-picking in Relion³¹. After cleaning by manual inspection and in 2D classification, per frame B-factor

weighting and translational movie alignment was applied to the remaining 282,000 particles. The colouring of the surfaces is according to the standard polymerase subunit colouring: A190, grey; A135, wheat; A49, light blue; A43, slate; AC40, red; A34.5, pink; Rpb5, magenta; Rpb6, silver-blue; AC19, yellow; Rpb8, green; A14, hot pink; A12.2, orange; Rpb10, blue; Rpb12, lemon. Template DNA, non-template DNA and RNA are depicted in medium blue, sky blue and red, respectively. The structures against greyed background indicate final EC and Pol I monomer structures.

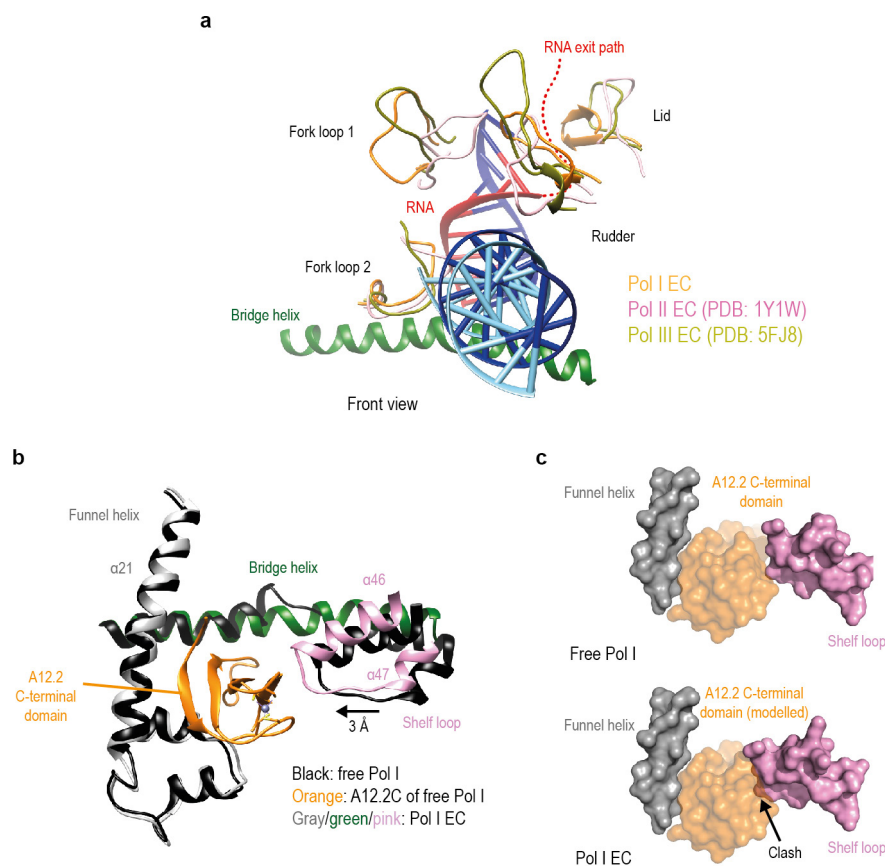


Extended Data Figure 3 | Quality of single-particle cryo-EM reconstructions. **a**, Top and bottom view of local resolution surface maps. **b**, Representative areas of the single-particle cryo-EM density for Pol I EC (left panel) and Pol I monomer (right panel). The A190 helix $\alpha 19$ (upper panel) and the A135 strand $\beta 40$ (lower panel) are depicted together with the refined model superimposed. **c**, Angular distribution of particle

images. Red dots indicate views with at least one particle assigned within 1° . Black shading represents the number of particles. The orientation occupancy is similar for both structures and covers most of the angles. **d**, FSC curves. Blue lines indicate the FSC between half maps of the respective reconstruction and red lines indicate FSC between the derived model against the single-particle cryo-EM map.

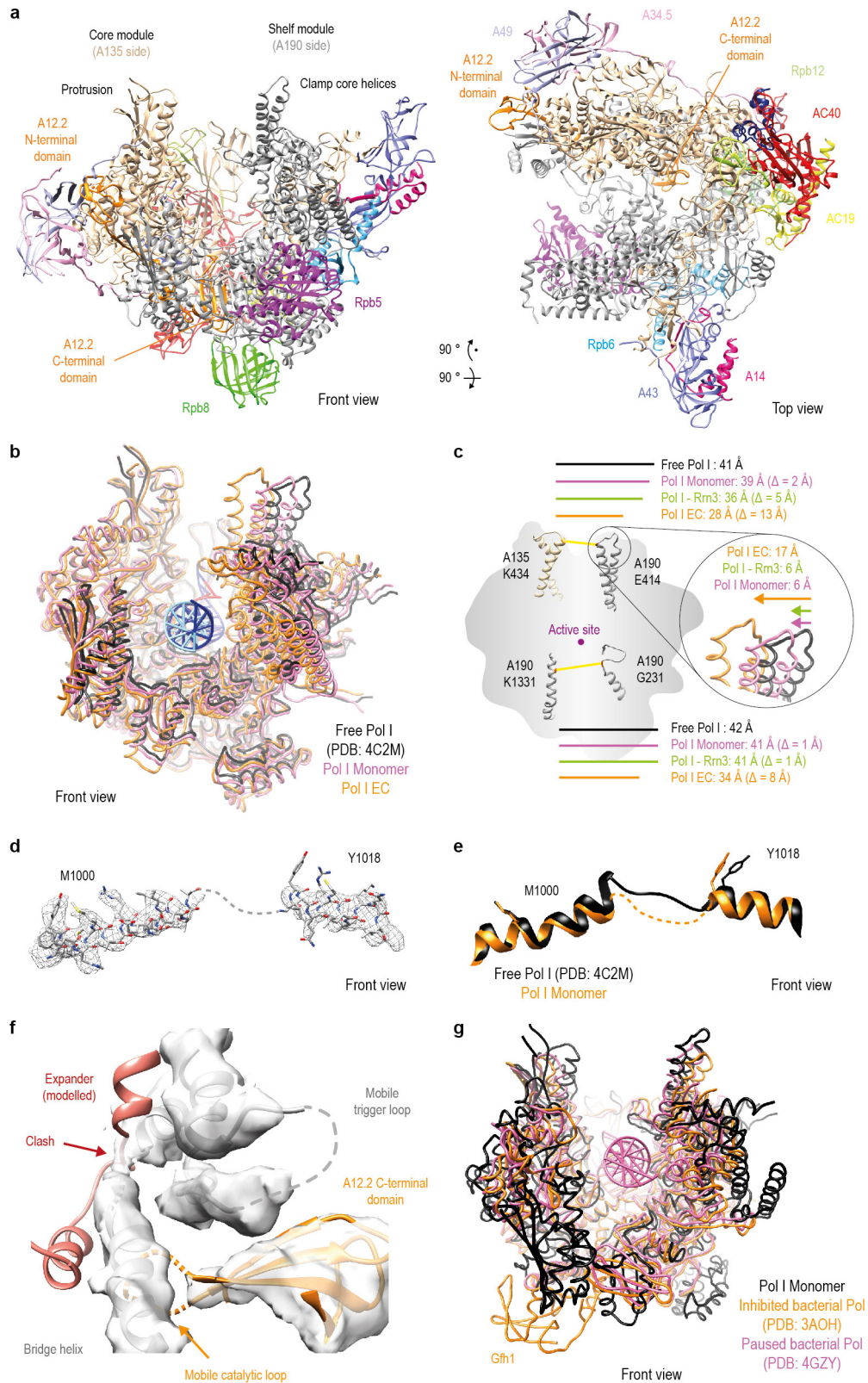


Extended Data Figure 4 | Comparison of all eukaryotic and the bacterial elongation complexes. All structures are depicted in front view. Bridge helix and active site are highlighted in green and magenta, respectively. Modules were defined as in Fig. 2d. **a**, Pol EC from this study. **b**, Pol II EC⁹. **c**, Pol III EC¹⁰. **d**, Bacterial EC¹⁵.



Extended Data Figure 5 | Additional details on Pol I EC. a, Cleft loops. Ribbon model of ECs of all three *S. cerevisiae* RNA polymerases superimposed on the bridge helix^{9,10}. Bridge helix (green) and downstream DNA together with DNA-RNA hybrid (blue and red) are given for Pol I. **b,** Ribbon model of free Pol I (PDB code 4C2M (ref. 5), black and orange) superimposed on the shown inner A190 funnel helix $\alpha 21$ with Pol I EC (grey, green and pink). As a consequence of cleft contraction, parts of the shelf module move in and reduce the width of the pore to impair binding

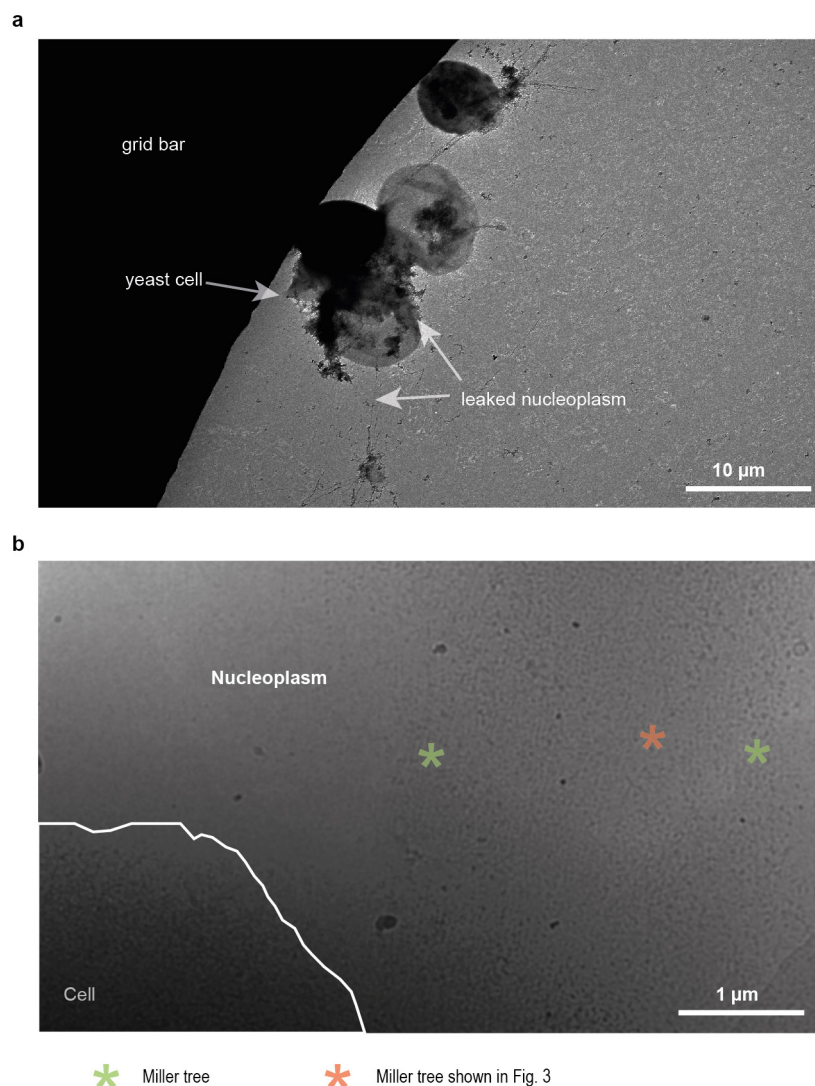
of the C-terminal domain of A12.2. **c,** Modelling the A12.2 C-terminal domain into the pore of the contracted Pol I EC results in a clash. In the upper part, a surface representation of domains in free Pol I shows that the C-terminal domain of A12.2 fills the pore that is lined by the A190 funnel helix $\alpha 21$ and loop 1572-1579 of the A190 cleft domain in the shelf module. In the lower part, cleft contraction observed in the EC reduces the width of the pore, causing a steric clash in the model.



Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | Free monomeric Pol I single-particle cryo-EM structure. **a**, Ribbon model of free, monomeric Pol I solved by single-particle cryo-EM. The views correspond to the 'front' and 'top' views with the incoming downstream DNA pointing towards the reader. Colour code as in Extended Data Fig. 2. **b**, Free Pol I (PDB 4C2M, black), Pol I Monomer (pink) and Pol I EC (orange) were superimposed onto A135. For clarity, only subunits A190 and A135 are shown. **c**, Scheme of the observed conformations as displayed in **b**. The cleft width was measured at two positions: (i) between residue K434 from chain A135 to residue E414 from chain A190 and (ii) between residue K1331 from chain A190 to residue G231 from chain A190. The distance bars above and below the polymerase cartoon indicate the distances between the protrusion and the clamp core helices (above) and at the entry site of downstream DNA (below). The difference (Δ) between the free Pol I (ref. 5) and the Pol I monomer, the Pol-I-Rrn3 (ref. 17) and Pol I EC, respectively, is given in

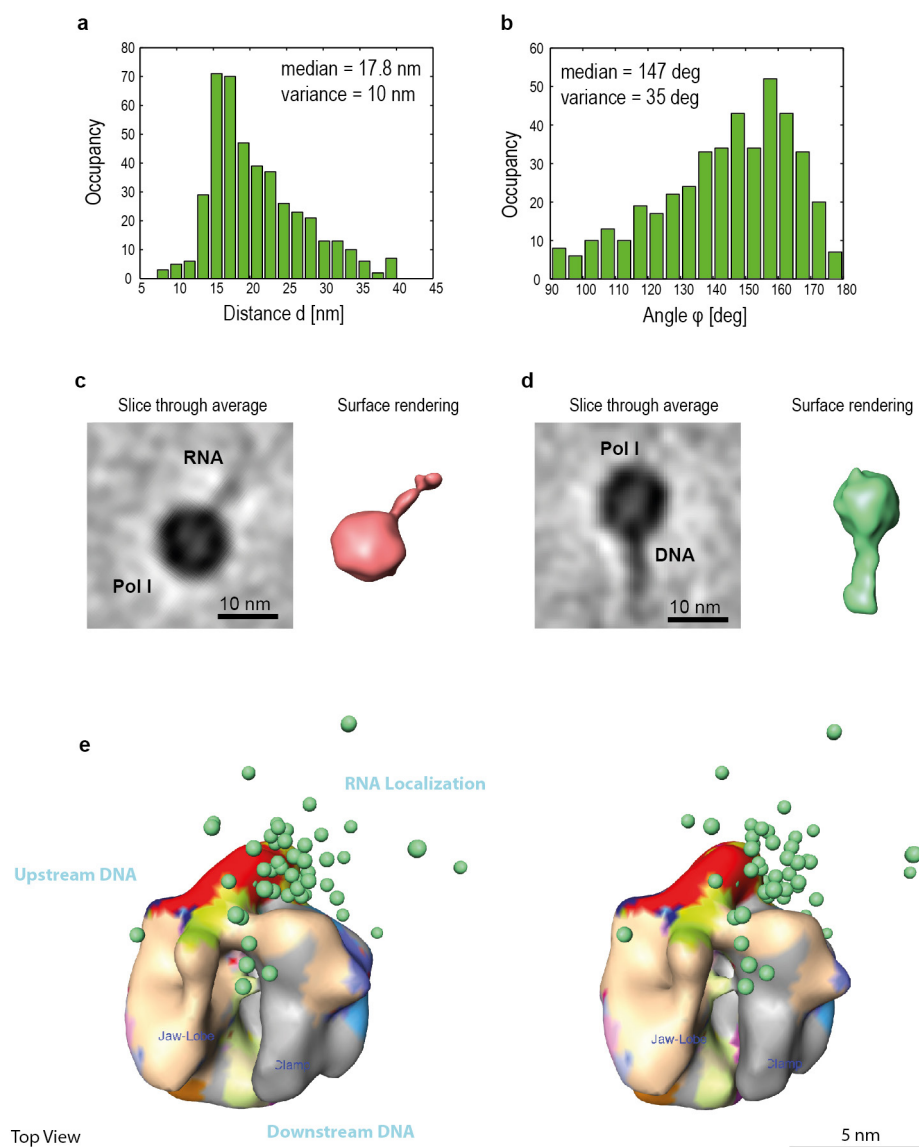
brackets. For the measurement of the relative movement of the clamp core helices shown in the magnified inset, all Pol I structures were aligned on the A135 subunit and the distance between residue E414 of subunit A190 of the free Pol I⁵ to the same residue of the Pol I EC, Pol I – Rrn3¹⁷ and Pol I monomer, respectively, was measured. **d**, Electron density of the bridge helix in the free Pol I Monomer. **e**, Comparison of bridge helices in the free Pol I monomer (orange) and free Pol I (black) from the crystal structure. **f**, Electron density (semi-transparent grey) is shown together with models for the bridge helix, trigger loop (both grey) and the C-terminal domain of A12.2. The expander (red) is not present in this structure but modelled here based on the crystal structure of the free Pol I dimer, revealing a clash. **g**, Inhibited²⁴ and paused²³ bacterial polymerase superimposed on Rpb1 of the free Pol I monomer. For clarity only A190 and A135 of Pol I are shown and the β' -NCD of the bacterial polymerases is excluded from the visualization.



Extended Data Figure 7 | Yeast cells, lysed to leak their nucleoplasm, prepared with negative stain and visualized under cryo conditions.

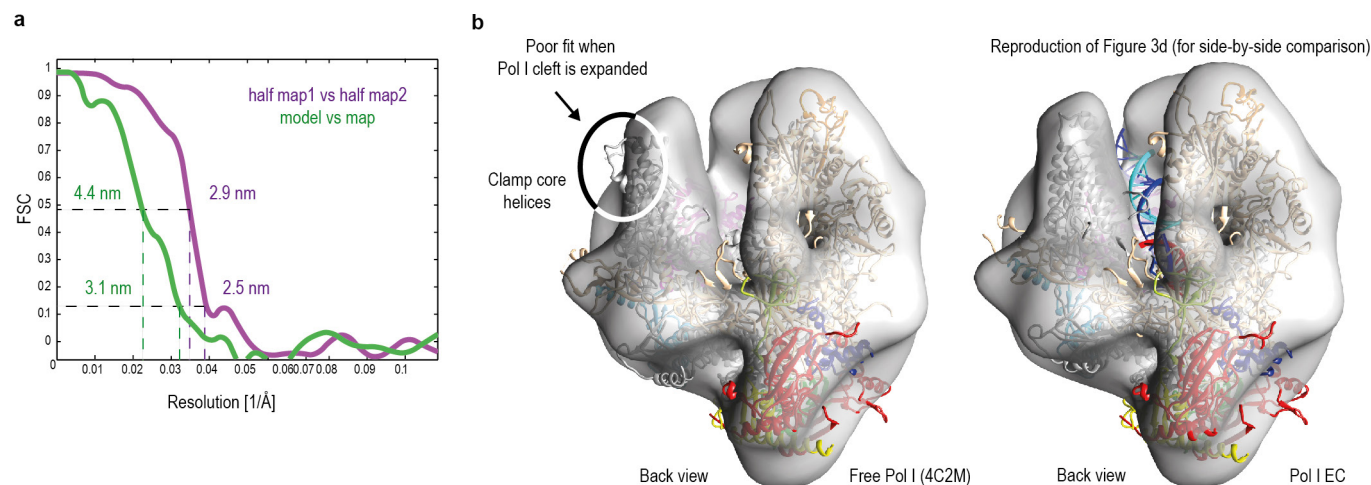
a, Electron micrograph of a negatively stained lysed yeast cell, with the nucleoplasm spread on the carbon support film. The upper left of the micrograph is occupied by the grid bar. The yeast cell has released the nuclear context on the grid, which appears as an electron-lucent

leakage. **b**, Electron micrograph of the spread nucleoplasm of a plunge-frozen yeast cell at close-to-native conditions. In the lower left corner, the remains of a yeast cell can be seen as an electron-dense patch. The nucleoplasm is embedded in an ice layer and the asterisks indicate three Miller trees found in the vicinity of this cell. The Miller tree indicated with the red asterisk was used for recording of the tilt-series in Fig. 3.



Extended Data Figure 8 | Relative positions of polymerases towards each other and of protruding nucleic acids. **a**, Histogram of centre-centre distance d of two consecutive Pols as depicted in (Fig. 3c). The reason for measuring distances smaller than 12 nm is that the enzyme is not completely spherical. Thus at certain rotational arrangements the centre-to-centre distance can be smaller than the average diameter of Pol I. **b**, Histogram of in-plane angle φ spanned by three consecutive Pols as depicted in Fig. 3c. **c**, Focused sub-tomogram averaging around the RNA. The RNA exits Pol I as an ~ 10 Å thick density, both in the slice and in the isosurface representation. **d**, Sub-tomogram average with the alignment

focused on the downstream DNA. The downstream DNA is a long, straight 2 nm density, both in the slice and in the isosurface representation. In both **c** and **d**, the Pol I molecule is a globular ~ 12 nm featureless density. **e**, Stereo pair of the sub-tomogram average shows the positions of the nascent RNA chain as green balls. The positions that were manually identified by three independent users, without previous knowledge of the positions of the sub-tomogram average, correspond closely to the position of the RNA exit site that was postulated by the X-ray crystallography structure⁵.



Extended Data Figure 9 | Comparisons between cryo-electron tomography and single-particle cryo-EM structures. **a**, FSC of the cryo-electron tomography structure with a resolution of 29 Å (purple line) and FSC between the cryo-electron tomography structure and single-particle cryo-EM structure with estimated resolutions of 44 Å and 31 Å, measured at FSC 0.5 and 0.143 criteria, respectively (green line). **b**, Poor fit of the expanded, free Pol I crystal structure⁵ (PDB: 4C2M) to the

cryo-electron tomography density (grey). In the expanded state a significant part of the clamp core helices are outside the cryo-electron tomography density (26% outside, 74% inside), while in the contracted state they almost completely enclosed (4% outside, 96% inside). In addition, in the expanded state 54% of the cryo-electron tomography density remains unoccupied compared to 19% of the contracted state (also compare Fig. 3d, reproduced here for comparison purposes).

Extended Data Table 1 | Model refinement statistics

	Pol I monomer	Pol I EC
Map CC (whole unit cell)	0.850 (0.844)	0.728 (0.724)
Map CC (around atoms)	0.703 (0.739)	0.733 (0.763)
rmsd (bonds)	0.008 (0.008)	0.007 (0.007)
rmsd (angles)	1.00 (0.86)	0.929 (0.864)
All-atom clashscore	32.11 (26.42)	13.72 (10.7)
Ramachandran plot		
outliers	0.4 % (0.2 %)	0.4 % (0.1 %)
allowed	4.9 % (4.6 %)	6.6 % (6.5 %)
favored	94.7 % (95.2 %)	93 % (93.4 %)
Rotamer outliers	1.0 % (0 %)	1.1 % (0 %)
C-beta deviations	1 (0)	1 (0)
EMRinger score	0.65 (0.72)	2.74 (2.93)
Molprobity score	2.56 (2.24)	2.12 (1.98)

Statistics for the core of Pol I excluding A14–A43 and A34.5–A49 are provided in parenthesis.

CAREERS

NATUREJOBS FACEBOOK Science-career posts, jobs, advice www.facebook.com/naturejobs

BLOG Stories and counsel for workplace support blogs.nature.com/naturejobs

NATUREJOBS For the latest career listings and advice www.naturejobs.com

ANDREW NATCHISON/IN PICTURES/GETTY



Scientists need to step forward if they are to ensure that politicians understand the importance of their work.

SCIENCE ADVOCACY

Get involved

Presenting science to politicians in a way they can understand can have good outcomes.

With so many science-based challenges facing the world, researchers who can help to inform and affect policy can have an outsized impact. We asked Connie Lee, Tamara Galloway and Niklas Höhne to describe how they have helped to shape government policy — and how others can learn from their experiences.

As chair of the public-policy committee for the American Society for Cell Biology (ASCB), Lee is a prominent advocate for science. She studied mammalian mitochondria before becoming an editor of *The EMBO Journal* and a deputy editor of *Cell*. As assistant dean for basic science at the University of Chicago in Illinois, she helps to oversee nine science departments.

After training as a physicist, Höhne turned

his attention to climate change, a field in which he hoped to make a global difference. As a founding partner of the New Climate Institute in Cologne, Germany, and a professor of greenhouse-gas mitigation at Wageningen University in the Netherlands, he works at the intersection of science and policy.

Galloway, an ecotoxicologist at Exeter University, UK, can say with certainty that her research — and her advocacy — have brought real-world results. Her testimony in front of Parliament in May helped to bring about a UK ban on microplastics in personal-care products, an important source of marine pollution. In June, she discussed her research on pollutants in front of a committee of the United Nations in New York City.



FIGHT FOR BASIC-SCIENCE FUNDING

Connie Lee, assistant dean for basic science at the University of Chicago in Illinois

Scientists have a lot of demands on their time. But getting involved in policy and advocacy is extremely important. Politicians hear from many lobbyists. If they don't hear from scientists too, we might be left out.

I got bitten by the policy bug in 2008 ►

► when I visited Capitol Hill, the home of the US Congress, as a representative of the ASCB. My dream is for every US scientist to visit Capitol Hill — you never know what questions politicians are going to ask. We met staff and elected members of Congress, and they had so many misunderstandings about science. A lot of people think that National Institutes of Health (NIH) funding only affects people in Bethesda, Maryland, where the NIH is headquartered. They don't realize that the funding spreads out to all 50 states, supporting research and creating jobs.

The lack of scientific understanding among policymakers can be frustrating. Members of the US House of Representatives will scan titles of NIH grants to find items that sound wasteful. These grants have been peer reviewed, but the politicians just look at the titles and take them out of context. It can be important work, but it's mocked and dismissed. We can't let that sort of thing get us down.

There's a communication gap between scientists and politicians. Scientists have to learn to explain the importance of their own work, whether they're talking to a policymaker, a dean or a potential donor. But we have to share a bigger message, too. We need to advocate for the institution of science and the importance of funding basic science. You never know where basic research can lead. The methodology behind CRISPR was discovered by looking at how bacteria protect themselves. Now it's used to edit genomes.

The ASCB lobbies for issues outside the lab, such as immigration and the importance of international collaboration. We want to make sure junior researchers from other countries receive visas that last long enough to allow them to get the training they need.

Policy and advocacy can take as much time as you're willing to give it. A lot of scientific societies have outreach positions, which is a great place to get started. You can join a government-relations board at your university or just offer a tour of your lab whenever a politician visits. And when you do get a grant funded, write to your local senator or representative and thank them for supporting science. It's baby steps, but we need to build relationships so they can see us as a resource.



THE FACTS MATTER

Niklas Höhne,
climate scientist
at Wageningen
University, the
Netherlands

Science covers the questions at the heart of society's problems. When it comes to climate change, it's absolutely essential that the

research community helps to translate science into options for policymakers.

I study international climate negotiations, such as the Paris agreement of 2015. The stated aim of the agreement was to limit warming to 1.5°C above pre-industrial levels. It requires much analysis to look at each country's emission proposals and then add them up to see whether they are on track to meet the overall goal. As I reported in November at the climate-change conference in Marrakesh, Morocco, our models show that some countries' current emissions proposals aren't sufficient to reach the Paris goal. Policymakers need this information so that they can adjust their country's emission targets, if they have the will to do so.

I would say that most governments are generally well-informed about climate change. The goal to limit warming to 1.5°C is stronger than the previous one of 2°C, and that's because politicians understood the evidence. Scientists were able to show that a 2°C rise wouldn't be safe for the planet.

Some politicians, including the president-elect of the United States, have denied that climate change exists. If individual politicians don't want to be convinced, there's not much more that scientists can do. Still, it's important to keep gathering data and reaching out to policymakers and the general public. The scientific community has a duty to continue to provide evidence and explain what we really know about human-caused emissions and global temperatures.

Every 6 years, for example, about 2,000 researchers work together to create a report for the Intergovernmental Panel on Climate Change on the current situation. It is a technical report that most politicians would have trouble understanding. But scientists can explain the key points and the take-home messages. Without that translation, their research won't have much of an impact.

In some parts of society, we seem to be moving to an era beyond factual argument. Emotion seems to matter more than the facts. We have total access to information, but we also have total access to misinformation. Scientists have to make the facts matter again. We can do that by communicating results in an accessible way.

I have been doing this for 20 years, and there have been a lot of setbacks. But I'm still hopeful. I got into climate change because I wanted to have an impact on the world, and I still think climate scientists can accomplish that. Despite the challenges, controlling climate change is doable. We have the technology we need to reach the goals. My main motivation is to give politicians the tools that they need to get this right. Progress may be slower than I hoped, but we'll see how things work out.



DELIVER YOUR MESSAGE

Tamara Galloway,
ecotoxicologist at
Exeter University, UK

The UK government is full of people who used to be bankers and lawyers. There's a great lack of scientific understanding. Most scientists aren't interested in becoming politicians, but it's still possible for them to become involved and inform policy.

As a scientist, I always wondered why it took government so long to act on issues, especially when the evidence was already clear. Then, in November 2015, I participated in the Royal Society's 'Week in Parliament' scheme, and spent a week in London shadowing a Member of Parliament (MP). It was an amazing experience, and it helped me to understand that government is a giant monolith. Change comes slowly — in most cases.

In May, I and two other environmental scientists addressed a parliamentary select committee of MPs on marine pollution caused by microplastics, spheres of plastic less than 5 mm in diameter. The committee had three hours to ask us any question they wanted, and we didn't know what to expect. I felt like I'd had been called into the headmaster's office.

The committee members asked intelligent and well-informed questions, trying to put everything in context. There's a lot of hysteria on the topic and websites with false information, so I needed to provide impartial scientific evidence. You don't want to sound as if you're pushing an agenda. I explained that microplastics, which are often found in cosmetics and shampoos, aren't actually toxic, but that they can disrupt the feeding and reproduction of many marine organisms.

Shortly after the hearing, the committee announced that microplastics will be banned from personal-care products in the United Kingdom by the end of 2017. The science we were doing had had a real impact, and I was amazed that it happened so quickly.

The pinnacle of my policy outreach — so far — is when I spoke about my research in front of a United Nations panel in New York in June. After that, I felt I could tackle anything.

The lesson for me is that we must speak up. Scientists tend to become more and more specialized, to the point where it can be difficult to talk to other researchers, let alone the general public. I use my children as a sounding board. If they understand, I know I'm ready. ■

INTERVIEWS BY CHRIS WOOLSTON

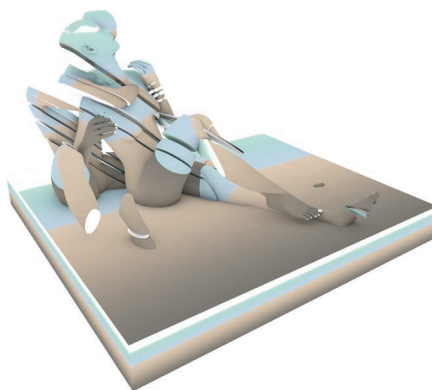
These interviews have been edited for clarity and length.

BREATHE THE LAST BITS OF AIR

Endgame.

BY EMILY MCCOSH

Are there stars outside?
Are you actually asking me that? Say it again.
Ask me again, if you're serious.
 I can ask again, I'm not afraid. Are there stars outside any more?
You should be afraid. And there may be.
 Well, go look.
How do you think I should go about doing that?
 Look out of the window, genius.
There are no windows ... there is no glass.
 There must be windows. How do you look outside without windows?
There're no windows. And there's no looking outside. Anyhow, if there were windows to look out of, there'd be nothing to look outside at.
 What kind of statement is that? I told you, look at the stars.
There are no stars. They've all winked out of existence. One. By. One...
 Utter rot.
Don't mock me.
 If there was anything left in the Universe worth mocking, it would be you.
One more insult, and I swear, I'll leave you all alone.
 You won't. You have nowhere to leave to. No path to walk upon. We're stitched together. Anyhow, how do you know the stars have gone out when there are no windows to look out of?
I've opened the door.
 Then open the door and look at the stars.
I'd rather not.
 I said go look at the stars! You never do as I ask. I can't even look for myself. My eyes are gone. They're whiter than milk —
There's no more milk.
 — they're larger than galaxies —
There are no more galaxies. No more stars. I told you that.
 — they're brighter than stars. They must be. Have you ever tried looking at my eyes while I'm sleeping? I think you must have. They must be brighter than stars.
 (She reaches out, lets her fingers hover over his eyes, but pulls her hand back.)
I've never looked at your eyes. Promise.
 Never?
Never. I've told you that. Listen. You're stubborn and you never listen to me. Why don't I leave you when you never listen to me?
 Well, maybe you love me?
Maybe? What kind of word is 'maybe'? I either love you or I don't.
 Well?
Well what?
 Do you?



Do I what?
 Love me?
Probably not.
 (There is silence like the end of the world.)
 Are you there?
I'm here.
 What were you saying?
I was saying I probably don't love you. I might have at one point but I probably don't now.
I wish I did, but I don't think there is such a thing as love any more. It caught the last train out of here. When all the people ran away to the dying stars, it went right along with them.
It's gone. It has died. Just like the Sun died.
 The Sun can't die. It's the Sun.
It's grey.
 Grey?
Grey. Less than black. More than white. Grey as slate.
 Is that what love is like now?
I think so. I can't be sure. I can't feel my heart.
 Can I feel it?
 (He reaches out, lets his hand hover over her chest. It doesn't come close enough to touch.)
I'd rather you didn't.
 (Withdraws his hand.)
I'd rather I did.
It doesn't matter. It's my heart. Even if I can't feel it.
 I could, you know? If I wanted to. If I really wanted to, you'd let me.
You couldn't. You can't even see me.
 Can you see me?
Not hardly.
 Well, I think you're right.
About what?
 About love like the dead Sun.
I probably am. It's a shameful thing to be right about.
 Is it?
 (He thinks, and laughs.)
 Maybe it's better to be right about it than to be wrong and think with all your heart that

you're right. If you could feel your heart, that is ... Are there stars outside?
There may be.
 Let's go see.
There's nothing left outside ... Anyhow, you can't walk.
 I know. You can carry me. Take me to the window. Let's go outside. We can use the door. You can carry me.
I can't.
 I know you can. You shouldn't lie.
Don't call me a liar. I'll leave you.
 You should leave me.
I should.
 Let's go outside.
 (They go outside, both of them.)
 Describe it to me.
Well, it's nothingness.
 Is that it?
It is. There's no description for nothingness. It's like sleeping, the world is sleeping ... Let me set you down.
 Yes ... Yes, do. Don't leave me though.
I won't.
 I know ... You can look at my eyes, if you'd like.
Will I like it?
 I don't know.
 (She crawls into his lap, opens his eyes and lifts his face.)
Well, what do you know? They are brighter than stars.
 Are they?
Yes, more beautiful too. But just as sad.
 Ah.
 (Kisses his lips, gently.)
I suppose ... you can feel my heart now ... if you'd like.
 I would ...
Then why hesitate?
 I don't know ...
 (He wraps his arms around her waist, lays his cheek against her breast.)
 Well, it's unsteady.
Yes, well, that has to do with you, not the world splintering to bits.
 Oh.
 (Sighs)
 Now?
 Now, what?
Now what do we do?
 Now? Breathe the last bits of air, I suppose.
 (Sighs)
 Breathe. ■

Emily McCosh writes and daydreams from her home in southern California. Visit her online at oceansinthesky.com or @wordweaveremily

ILLUSTRATION BY JACEY